INTERACTIONS^3: LANGUAGE, DEMOGRAPHICS, AND PERSONALITY; AN IN-DEPTH ANALYSIS OF GERMAN TWEETS

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

BY

MATTHIAS LUDWIG RAESS

DISSERTATION ADVISOR: DR. CAROLYN MACKAY

BALL STATE UNIVERSITY

MUNCIE, INDIANA

MAY 2018

INTERACTIONS^3: LANGUAGE, DEMOGRAPHICS, AND PERSONALITY;

AN IN-DEPTH ANALYSIS OF GERMAN TWEETS

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

BY

MATTHIAS LUDWIG RAESS

APPROVED BY:

Committee Chairperson	Date
Committee Member	Date
Committee Member	Date
Committee Member	Date
Dean of Graduate School	Date

BALL STATE UNIVERSITY
MUNCIE, INDIANA
MAY 2018

ACKNOWLEDGEMENTS

Overcoming challenges throughout an intellectually taxing endeavor, such as writing a dissertation, certainly required me to self-reflect, and to adapt to new trials frequently. It also challenged me to keep an open mind in the face of opposing views and findings. While this dissertation was an exercise in self-discipline, the entire process was an enjoyable journey, during which I tremendously improved and honed my data science skills, my statistical and programming skillset in R, my linguistics skills, and, with deadlines always looming, also my time and project management skills.

I am delighted to report that this dissertation was completed in its entirety without the help of any form of higher being or deity, imagined or otherwise. However, it goes without saying that the completion of a project of this size is, directly and indirectly, the result of the help and support of many very real people.

Accordingly, I would like to express my gratitude to my wonderful advisor and committee chair, Dr. Carolyn MacKay, who encouraged this project from the beginning with unwavering support. She also played an integral part in forging me into the researcher I am today, by not only critically challenging my ideas, but also making many valuable and detailed suggestions for improvements along the way. Carolyn's motivation, interest, confidence in my capabilities, and shared excitement helped tremendously throughout the thorny road to completion. I'm also very grateful for the support of my other incredible committee members. Dr. Elizabeth Riddle, who read my dissertation drafts with painstaking attention to detail, providing valuable feedback along the way, and Dr. Frank Trechsel for engaging with me in insightful and inspiring conversations throughout the years and for critically challenging my project, especially during the proposal stage. I would also like to express my gratitude and appreciation to Dr. Holmes Finch, who not only taught me the intricacies of multivariate statistics, but also played an integral role in vastly expanding my overall statistics skills. Through many great conversations, Holmes provided critical feedback to statistical analyses and methods and helped me navigate the minefield that is frequentist statistical analysis.

I would also like to express my gratitude to Dr. James Pennebaker (UT Austin) for having a personal sit-down with me to talk about the intricacies of LIWC and Dr. Markus Wolf (University of Zurich, former University of Heidelberg), co-author of the German LIWC dictionary, for many insightful comments and hints pertaining to LIWC usage and analysis. In addition, I would like to thank Meik Michalke, author or the koRpus-R-package, for his help, Dr. Bodo Winter (University of Birmingham) for critical comments, suggestions, and feedback, especially during the developmental stages of this project, and Dr. Tatjana Scheffler (University of Potsdam) for providing insightful information on German Twitter. I would also like to thank Prof. Dr. Volker Depkat (University of Regensburg), Dr. Alexander Kautzsch (University of Regensburg), and Prof. Dr. Thomas Hoffmann (University of Osnabrück) for helping to pave the way to this endeavor with their recommendation letters.

To Wendy, thank you for keeping me sane throughout this research roller-coaster, always lending an ear when I needed to vent or complain about things, for being understanding of my needs, the obstacles I faced along the way, and for never doubting me. To my best friend, Philipp, thank you for your wit, intellect, and inquisitiveness, and for challenging my ideas, providing critical feedback, and thought-provoking suggestions from a non-linguist angle. Thank you also for your continued support, and for always being there for me. To all my other friends, near and far, especially Maximilian H., Maximilian Z., Wendelin, Christian, Matt, and Gene, thank you for encouraging words, support, and great conversations. Thank you also for being there, keeping in touch, and helping me through this journey. Last but certainly not least, thank you to my family, for always encouraging my endeavors however crazy they may be. To my sisters, Veronika and Barbara, thank you for being there across the pond, lending a shoulder to lean on when I was frustrated and for many happy moments when I travelled home. You're the best sisters a person could ask for.

I accept responsibility for all the ideas in this dissertation. All remaining errors, gaps, and omissions are my own.

***

I would also like to acknowledge Ball State University for funding this research and the preceding studies through a graduate assistantship with the Intensive English Institute and various research assistantships, as well as the Honor Society of Phi Kappa Phi for funding through a Love of Learning Award. Further, I would like to acknowledge the Intensive English Institute, the Ball State English Department, and the BSU Aspire Internal Grant Program for financial support for conference travel throughout the years.

DEDICATION

There are only two lasting bequests we can hope to give our children. One of these is roots, the other, wings. – Johann Wolfgang von Goethe

Zwei Dinge sollen Kinder von ihren Eltern bekommen: Wurzeln und Flügel. – Johann Wolfgang von Goethe

I would like to express my sincerest gratitude and appreciation to my parents, not only for giving me both roots and wings, but also for making this endeavor possible. I remain forever in your debt and I know that this dissertation, and the preceding studies would not have been possible without your numerous sacrifices, continued and unwavering support, understanding, and personal investment in my success.

I ACCORDINGLY DEDICATE THIS DISSERTATION, THE PINNACLE OF MY INTELLECTUAL ACCOMPLISHMENTS, TO MY PARENTS

LUDWIG AND GERTRUD RAESS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

DISSERTATION: Interactions^3: Language, Demographics, and Personality; an In-Depth Analysis of German Tweets

STUDENT: Matthias Ludwig Raess

DEGREE: Doctor of Philosophy

COLLEGE: Sciences and Humanities

DATE: May 2018

PAGES: 260

This study examines the interactions between German Twitter users' personalities, the specific features of their tweets, including the emoji density, types of hashtags and their density, and the percentage of various LIWC word categories, such as emotion words (e.g. positive/negative), and how these variables interact with gender. This tripartite analysis in conjunction with questionnaire data and online Twitter data not only advances our understanding of how gendered-language, and thus perceived stereotypes, and personality in conjunction with linguistic cues behaves on Twitter, but also how an out-of-the-lab sample contributes to the generalizability of results and insights gleaned from these analyses. The significance, therefore, not only lies in the combination of research areas (linguistics and psychology) and the analysis approach, but also the fact that Twitter studies focusing on German, incorporating personality measures are far and few between.

The broad research design of this study is quantitative and interdisciplinary in nature, encompassing both linguistics and psychology. Participants, $N = 62$, filled out an online questionnaire providing demographic information and information on their personality traits through the German short version of the Big Five Inventory, the BFI-10 (Gosling, Rentfrow, &

Swann, 2003; Rammstedt & John, 2007; Rammstedt, Kemper, Klein, Beierlein, & Kovaleva, 2012). In addition, participants' tweets, $N = 19{,}772$, were collected using the Twitter API and then combined with their demographic information, including their Big5 scores. The tweets were then analyzed with the software, Linguistic Inquiry and Word Count (LIWC) (Pennebaker & King, 1999), which made it possible to quantify linguistic features, such as percentage of emotion words, or anger words, for example. In addition, quantitative measures of users' frequencies of hashtags, including a hand-coded hashtag subset, $n = 2{,}666$, and emojis, including sentiment scores were used for quantitative analyses.

The study furnishes new confirmatory evidence for previous findings regarding significant positive correlations between positive feeling words and extraversion, agreeableness, and neuroticism. A significant positive correlation between neuroticism and anxiety words was also confirmed. In addition, extraversion turned out to be a significant predictor for sentiment scores on Twitter, indicating that extroverts benefit more from being active online. In terms of LIWC categories, gender was a significant predictor for both positive emotion words, and positive feeling words, with females using more in both categories. Gender also turned out to be a significant predictor for anger words, swear words, occupation words, and words related to money, with women using higher percentages in these categories, which contradicts previous research in an English-language context, and contrary to my own expectations.

The study thus offers new insights into the differences in relationship to gender and context-dependent language use, adding support for some of the key arguments. Specifically, female German Twitter users turned out to use language differently compared to previous findings, e.g. lower percentage of words related to tentativtiy. In addition, German female Twitter users seem to use the social medium for different, more professional purposes. The study

thus addresses contentious previous findings by adding new information to the perceptions of gender-dichotomous language use, indicating that it does not necessarily follow the same patterns across genres, i.e. different social media, prompting a re-thinking of some previous findings. The statistical significance of the findings allows us to make conservative and careful generalizations to the larger German Twitter user base.

Keywords: LIWC, sociolinguistics, Twitter, German, Big Five, language and gender

CHAPTER ONE: INTRODUCTION

The saddest aspect of life right now is that
it gathers knowledge faster than society gathers wisdom. - Isaac Asimov

Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less. - Marie Curie

1.1 Introduction

It has been well established that individual personality traits are closely linked to a person's use of language (Klages, 1926). Allport and Odbert's (1936) finding that language encodes individual differences in humans laid the groundwork for long-standing research into the relationship of personality traits with language use and linguistic cues. Researchers such as Gottschalk and Gleser (1969) and Weintraub (1989) built on these early studies by investigating how psychological states can be assessed through content analysis and how verbal behavior relates to personality. While recent studies have addressed these issues with offline language data (e.g. Hirsh & Peterson, 2009; Weisberg, DeYoung, & Hirsh, 2011), only a few studies have looked into how personality and language are related on social media (e.g. Back et al., 2010; Qui, Lin, Ramsay, & Yang, 2012). Moreover, Twitter has been left out of the equation as its own hybrid-genre almost altogether.

It has also been well established that perceived gender stereotypes about language use abound. While considered highly contextual by many sociolinguists, recent studies that revolve around automated gender prediction from text have tied some of these perceived stereotypes to the word level: For example, females have been shown to use more modal verbs and other words expressing tentativity (Biber, 1989; Mulac, Bradac, & Gibbons, 2001; Newman, Groom, Handelman, & Pennebaker, 2008), which can make any statement more indirect. Women are also said to discuss people more and communicate internal processes, thoughts, emotions, senses, and use more past and present tense verbs. Overall, women are perceived to use more emotional language than men, which can make their language more involved, i.e. higher use of pronouns, adverbs, and verbs, and, thus a preference for topics revolving around people, relationships, and internal states can be attributed to those stereotypes (Argamon, Koppel, Fine, & Shimoni, 2003; Aries & Fern, 1983). Men, on the other hand, are said to use more swear words, and longer words. They discuss external events, objects and processes pertaining to occupation, money, and sports more, which is evidenced in a higher use of numbers, prepositions, and longer words — their language is less emotional overall and their style is more informational, i.e. more nouns, adjectives, and prepositions. Men's topics stereotypically revolve around objects such as cars and computers, and political events (Coates, 1993; Johnson, 1994).

Further, research has established a connection between Twitter usage behavior and the language that is being used. The importance of users' languages was confirmed in one of the first studies on this issue by Hong, Convertino, and Chi (2011), who found that users do indeed use Twitter, and different languages on it, for different purposes with varying online behavior. For example, Germans tend to use hashtags quite a bit with, 18% of German tweets containing hashtags, as compared to only 5% in Japanese tweets (Hong et al., 2011). Overall, Germans, on average, seem to be more likely to include URLs and hashtags than users of other languages, which could be indicative of more content-related tweet behavior (Hong et al., 2011). This could also indicate user behavior that diverges from the canonically intended use of the hashtag as a tag to create a meta-datum, i.e. a tag hashtag, versus one that adds a comment or additional information tag, i.e. a commentary hashtag (Shapp, 2014; Twitter Inc., 2017g).

According to Statista.com (2016), English (34%), Japanese (16%), Spanish (12%), Malay (8%), and Portuguese (6%) make up the top-five Twitter languages. What has not changed is the fact that German does not appear in the top five, not even makes it into the top ten (Hong et al., 2011; Scheffler, 2014). It is time to answer the call for further research into German as a Twitter language (Hong et al., 2011; Scheffler, 2014; Weerkamp, Carter, & Tsagkias, 2011).

To achieve this goal, I implemented a two-fold data-collection approach consisting of a questionnaire in conjunction with Twitter data. Sixty-Two German participants, residing in Germany, filled out a questionnaire on demographic information, including the German short version of the Big Five trait assessment (BFI-10). Based on the lexical hypothesis (Allport & Odbert, 1936), psychologists developed what is now known as the Big Five personality inventory (John, Donahue, & Kentle, 1991). The Big Five now comprise conscientiousness, agreeableness, neuroticism, openness, and extraversion (Goldberg, 1992; John & Srivastava, 2001). Although the entire inventory comprises 240 questions (NEO-PI-R), short 10-item versions have also been developed for the US (TIPI) and Germany (BFI-10) (Gosling et al., 2003; Rammstedt & John, 2007). Today, the Big Five inventory is the gold standard for assessing personality traits (Mairesse, Walker, Mehl, & Moore, 2007), with the model's general consistency across age, gender, and cultural lines (John, 1990; R. McCrae & Costa, 1990), as well as its validity across different languages having been established (Digman, 1990; John, 1990; R. McCrae & John, 1992; R. R. McCrae, 1989). This allowed me to investigate interactions between participants' demographic information and personalities, i.e. their measurements on the Big Five traits.

In addition, every participant's tweets were collected via Twitter's open API.[1] Twitter's limit for the collection of individual user's tweets is 3,240. Thus, a unique situation, in which I could analyze Twitter data beyond previous limitations, by bridging the gap between missing or sparse demographic information and users' tweets, was created. This way, demographic information could be directly and precisely linked to users' tweet behavior and their personality profiles, which made it possible to analyze the data quantitatively. In addition, linking participants' questionnaires to their respective Twitter accounts, allowed me to circumvent the problem of not having geo-tagged tweets, which were limited to Germany, and not all German speaking countries. Most importantly, by collecting participants' tweets, I gained access to users' natural language on Twitter. While most tweets today are written in English, estimates now project that half of all daily tweets are not English, and only thirty percent of tweets originate in the U.S., with Japanese and Portuguese tweets coming in second and third respectively (Semiocast, 2010, 2010, February 24). Scheffler (2014) pointed out that only roughly one percent of tweets are written in German, including tweets from Austria and Switzerland.

The question that arises is how we even begin to analyze large amounts of language data and their linguistic features. Utilizing the word-count approach (Pennebaker, Booth, & Francis, 2007; Pennebaker & King, 1999), Pennebaker and King (1999) built the software tool, Linguistic Inquiry and Word Count (LIWC), to tackle this very problem. Mehl (2006) referred to LIWC as currently being the most-used, and best-validated software in psychological research for automated text analysis. As a dictionary based analysis tool, LIWC utilizes a dictionary (~6,400 words in the 2015 version) with different categories (74) as a baseline, against which text input is compared and percentages for individual word and content categories are computed. With

[1] Twitter's application programming interface (API) lets developers or users automatically access Twitter to collect tweets, post updates, etc. via a registered application (Twitter Inc., 2017c).

correlations ranging from .20 to .81., Pennebaker and Francis (1996) (also see Back et al., 2010; Golbeck, Robles, Edmondson, & Turner, 2011; Golbeck, Robles, & Turner, 2011; Mairesse et al., 2007; Pennebaker & King, 1999; Qui et al., 2012; Schwartz et al., 2013; Tausczik & Pennebaker, 2010; Vazire & Mehl, 2008; M. Wolf et al., 2008; Yarkoni, 2010) demonstrated the external validity of word categories to be included in LIWC. Since the software is dictionary based, usage in a language other than English requires a dictionary in said language, which has to be tested for equivalence to the English dictionary and for robustness in terms of errors. In a twofold approach, Wolf et al. (2008) did just that: they tested the equivalence of a German dictionary for LIWC and the robustness of that dictionary in terms of spelling errors (the German 2001 LIWC-dictionary contains 7,598 words). Since the data set in this study is relatively large, LIWC was used to get the percentages for LIWC word categories, such as emotion words, words pertaining to occupation, social processes, and overall word frequencies.

1.2 Rationale of the Study

Most research on Twitter has been on English, and predominantly U.S. English, with German as a social media language remaining understudied. Furthermore, much high-caliber sociolinguistic research comes out of North America (among others, Eckert, 2011a, 2012; Labov, 1972, 1990; Tagliamonte, 2006a; Tagliamonte, 2014; Tagliamonte & D'Arcy, 2007; Tannen, 1990a, 1990b), which makes many investigations rather North-America-centric. Thus there is a need for further sociolinguistic research in German, and in particular, research using up-to-date, social media data from a representative, off-campus, out-of-the-lab participant sample.

1.3 Significance of the Study

Even though there are now numerous studies revolving around the issue of gendered-language use on social media, and the interactions of gender and LIWC word categories, German has only received a small amount of scholarly attention and continues to be neglected. Scheffler (2014) has offered a snapshot of German Twitter, but her study barely scratches the surface of what still warrants investigation. Further, LIWC has not been used extensively, if at all, in a German language environment, and especially for interdisciplinary research that bridges the gap between linguistics and psychology. In addition, while previous studies have revolved around similar questions pertaining to social media and language use in conjunction with demographic variables, few use the same social medium, which reduces their comparability, if only to some extent.

There is long-standing research on personality, and personality linked to online behavior, especially using the Big Five inventory (e.g. Qui et al., 2012; Schwartz et al., 2013), as well as extensive research on sociological variables, such as gender, race, age, and education in relationship to language (e.g. Eckert, 2011a; Holmes, 2001). There are also a few studies that link personality to language use and linguistic cues (e.g. Pennebaker, Mehl, & Niederhoffer, 2003; Tausczik & Pennebaker, 2010), and research that links personality traits to sociological variables (e.g. Weisberg et al., 2011). Finally, there is no lack of research on English tweets in general (e.g. Rehbein & Ruppendorfer, 2013). However, at this time, to the best of my knowledge, no study has been carried out that focuses on the tripartite interaction(s) between language (linguistic cues), sociological/demographic data, personality, with these variables being looked at against a German Twitter backdrop.

Accordingly, this study answers the call for further research into Twitter related linguistic research in general, and research on German in particular (e.g. Hong et al., 2011; Scheffler, 2014; Weerkamp et al., 2011). By using questionnaires, it includes more reliable and accurate demographic information, compared to studies, which used gender prediction algorithms (e.g. Deitrick et al., 2012; Kokkos & Tzouramanis, 2014; Miller, Dickinson, & Hu, 2012), for example. As a result, the limitations of using a sample of tweets as the only data source are surpassed by linking demographic and personality information to it. Further significance lies in this study's methodology. Using snowball sampling, thus going beyond convenience sampling in a campus environment used in previous studies, the data come from a relatively representative sample of participants, including representative natural language data with good geographical spread across Germany, extending similar prior work, some of which exclusively relied on lab-based data. In addition, I used generalized additive models, where appropriate, to reveal non-linear relationships between predictor and outcome variables. GAMs are still somewhat underused as statistical analysis tools. Thus, this study also contributes to the area of frequentist applied statistical research.

Overall, the number of linguistic studies investigating social media is still relatively low, compared to other genres and registers. Since there is only little previous research on Twitter in Germany, studying this genre and register fills in gaps in this area and provides a new analysis for comparison with other researchers' findings.

1.4 Operationalizations of Important Key Terms

1.4.1 Word Categories from the 2001 German LIWC Dictionary

Linguistic Inquiry and Word Count (LIWC) uses a dictionary based approach to calculate percentages for different word categories. LWIC comprises 74 word categories, some of which encompass emotions, another includes swear words, and still another includes words related to occupation, for example. The specific word categories used in this study are presented below. The German 2001-LIWC dictionary currently comprises 7,598 words. Thus, the words listed here for expository purposes are only a subset of the words included in each category.

Emotion words. Emotion words are generally subsumed under the category of affect words, which encompass positive emotion words, negative emotion words, and words related to anxiety, anger, and sadness. Semantically, all of these words invoke imagery that is highly related to a person's mental state, and how they feel. Just like many other words, affect words are understood along semantic axes, e.g. "good-bad," or "horrible-glorious" (Grefenstette, Qu, Evans, & Shanahan, 2008). In linguistics, this "linguistic scale" was defined by Levinson (1983) as a set of alternate or contrastive expressions that can be arranged on an axis by degree of semantic strength, which relates to the idea of semantic fields (Berlin & Kay, 1969; Lehrer, 1974). For example, words such as *good, nice, excellent, positive, fortunate, correct,* and *superior* fall into the positive emotion words category, while words such as *bad, nasty, poor, negative, unfortunate, wrong*, and *inferior* fall into negative emotion words category (Grefenstette et al., 2008).

Positive emotion words. The category of positive emotion words from the German LIWC dictionary includes, but is not limited to, the following: *aktiv* 'active,' *angenehm* 'pleasurable,' *angeregt* 'aroused,' *anhänglich* 'attached,' *anlächeln* 'smile at,' *beeindruckt* 'impressed,' and *beliebt* 'popular.'

Positive feeling words. The category of positive feeling words from the German LIWC dictionary includes, but is not limited to, the following: *Freude* 'joy,' *fröhlich* 'happy,' *Gefühle* 'emotions,' *lächeln* 'smile,' *Leidenschaft* 'passion,' *Liebe* 'love,' *Wohlwollen* 'benevolence.' Due to their semantic similarity, both positive emotion words and positive feeling words were conflated into a single category in newer English versions of LIWC, starting with the 2007 version (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007, p. 8). Markus Wolf, one of the co-authors of the German LIWC dictionary, confirmed that there are no substantial differences between both categories, which is why they inherently correlate highly with each other. Positive feeling is a subcategory of positive emotion, which means that almost all positive feeling words are contained in the positive emotion category (Wolf, M., personal communication, September 11, 2017).

Negative emotion words. The category of negative emotion words from the German LIWC dictionary includes, but is not limited to, the following: *bedroht* 'threatened,' *bedrückend* 'depressing,' *benachteiligt* 'disadvantaged,' *Beschwerde* 'complaint,' *besorgt* 'worried,' and *eingeschüchtert* 'intimidated.'

Anger words. The category of anger words from the German LIWC dictionary includes, but is not limited to, the following: *Beschwerde* 'complaint,' *böse* 'evil,' *eingeschüchtert* 'intimidated,' *gedroht* 'threatened,' *habgier* 'greed,' *missbrauchen* 'abuse,' *Rache* 'revenge.'

Swear words. Swear words belong to a class of taboo words commonly avoided in formal speech. The category of swear words from the German LIWC dictionary includes, but is not limited to, limited to, the following: *abschaum* ‘scum’, *fuck*, *ärsche* ‘asses,’ *bastard* ‘bastard,’ *bauerntölpel* ‘redneck,’ *bekloppt* ‘retarded,’ *bescheuert* ‘crazy,’ and *depp* ‘idiot.’

Modal verbs and words that express tentativeness. The category of modal verbs and other words that express tentativeness from the German LIWC dictionary includes, but is not limited to, the following: *könnte* ‘could,’ *sollte* ‘should,’ *würde* ‘would,’ *vielleicht* ‘maybe,’ *vorsichtig* ‘careful,’ and *wahrscheinlich* ‘probably’ (see Appendix C, p. 258 for full list).

Social concerns. The category of words pertaining to social concerns from the German LIWC dictionary includes, but is not limited to, the following: *Vorschlag* ‘suggestion,’ *Vorwurf* ‘accusation,’ *widersprechen* ‘contradict,’ *zugeben* ‘admit,’ and *zulassen* ‘allow.’

Family and friends. The category of words pertaining to family and friends from the German LIWC dictionary includes, but is not limited to, the following: *Bruder* ‘brother,’ *Ehefrau* ‘wife,’ *Ehemann* ‘husband,’ *Eltern* ‘parents,’ *Familie* ‘family,’ *Geschwister* ‘siblings,’ *Gesellschaft* ‘society,’ *Kollege* ‘colleague,’ *Freundin* ‘girlfriend,’ and *Gast* ‘guest.’

Occupation. The category of words pertaining to occupation from the German LIWC dictionary includes, but is not limited to, the following: g*efeuert* ‘fired,’ *Gehalt* ‘salary,’ *Gehalterhöhung* ‘raise,’ *Geschäft* ‘business,’ *Gewinner* ‘winner,’ and *Herausforderung* ‘challenge.’

Job. The category of words pertaining to job from the German LIWC dictionary includes, but is not limited to, the following: *Abfindung* ‘settlement,’ *Abteilung* ‘department,’ *Beruf* ‘job,’ *Besitz* ‘possessions,’ *Chef* ‘boss,’ and *Fähigkeit* ‘skills.’ Like the positive emotion and positive feeling categories above, the occupation and job categories are semantically very similar with extensive overlap between the categories. Thus, they were also conflated into a new work-category starting with the 2007 English version of LIWC (Pennebaker, Chung, et al., 2007, pp. 5-12).

Achievements. The category of words pertaining to achievements from the German LIWC dictionary includes, but is not limited to, the following: *Fleiß* ‘diligence,’ *fortgeschritten* ‘advanced,’ *geleistet* ‘achieved,’ and *Kämpfer* ‘fighter.’

Sports. The category of words pertaining to sports from the German LIWC dictionary includes, but is not limited to, the following: *Lauf* ‘run,’ *Mannschaft* ‘team,’ *Spiel* ‘game,’ *Leistung* ‘performance,’ and *Workout* ‘work out.’

Money. The category of words pertaining to money from the German LIWC dictionary includes, but is not limited to, the following: *Aktie* ‘stock,’ *Lohn* ‘wages,’ *Armut* ‘poverty,’ *bezahlen* ‘pay,’ *leihen* ‘borrow,’ *Darlehen* ‘loan,’ *Ersparnisse* ‘savings,’ and *Geiz* ‘stinginess.’

1.4.2 Sentiment Scores

Sentiment scores are based on a -1, 0, +1 range following Novak, Smailović, Sluban, and Mozetič (2015b). Calculating sentiment scores for 751 emojis, Novak et al. (2015b) found a mean sentiment score of +0.365 for tweets containing emojis. In this study, the sentiment score per tweet is based on the emojis present in a given tweet, using Novak et al.’s (2015b) sentiment scores to match the emoji. If there are two or more emojis present in a single tweet, the sentiment score for this tweet is the mean of those emojis’ sentiment scores.

1.4.3 Big Five Scores

The Big Five factor model comprises *Extraversion* ‘extraversion,’ *Verträglichkeit* ‘agreeableness,’ *Gewissenhaftigkeit* ‘conscientiousness,’ *Neurotizismus* ‘neuroticism,’ and *Offenheit* ‘openness.’ These factors are usually measured with Likert-scales, ranging from one (strongly disagree) to five (strongly agree). Participants’ Likert-scale answers to statements (ten for the BFI-10) for each factor were averaged to obtain individual scores for each factor. Jointly, the resulting five scores represent an individual's personality (John & Srivastava, 2001). A high score on extraversion means that an individual is sociable, outgoing, talkative, and assertive, while a high score on agreeableness indicates that an individual is cooperative, helpful, and nurturing. People scoring high on conscientiousness are responsible, organized, hard-working, and reliable. Individuals, who have high scores on the neuroticism trait are anxious, insecure, and sensitive, while people, who score high on openness are curious, intelligent, and imaginative (Golbeck, Robles, & Turner, 2011).

1.4.4 Hashtag Types

Tag hashtags are canonically used to tag topics, which is done by naming a concrete entity such as a person (#Obama, #Hillary), a place (#Paris, #TheHaven), a company (#Apple, #AMC), or an event (#Thanksgiving, #SummerOlympics). While these tags can be directed at different levels of the public ranging from tags relevant to tweets for the general public all over the world, or country, they also tag entities that are only relevant to certain individuals or a select group of Twitter users/followers. For example, #Paris is a hashtag that speaks to a large audience while #BSU has a much more focused local use and addresses students, faculty, alumni, and staff at Ball State University (Shapp, 2014).

Commentary hashtags serve to "add additional meaning to the main semantic content of the tweet, and are not intended to practically connect the tweet to others that use the same hashtag" (Shapp, 2014, p. 7). Usually, commentary hashtags add an evaluation to what the author of the tweet just said. Routinely, this process adopts the following syntactic pattern: "Text body of the tweet in a sentence. #evaluation." For example, "When someone tells u its not safe to travel to a foreign place alone just cause ur female. #ridiculous #wearenotincapable."

1.5 Hypotheses

This study seeks to show how interactions between language, gender, and personality interact in German tweets. As the data-collection approach is twofold and includes Twitter data in addition to questionnaires with demographic information and personality profiles, the following hypotheses, which fall into four categories were tested: personality and linguistic features, Twitter measures in relationship to gender, gender effects and LIWC categories, and word-based measures as related to gender, were tested.

1.5.1 Personality and Linguistic Features

(1a) There will be a significant positive correlation between an extraverted personality (as measured by the score for extraversion in the Big Five factor model) and the percentage of positive emotion words.

(1b) There will be a significant positive correlation between an agreeable personality (as measured by the score for openness in the Big Five factor model) and the percentage of positive emotion words.

(1c) There will be a significant negative correlation between an agreeable personality (as measured by the score for openness in the Big Five factor model) and the percentage of swear words.

(1d) There will be a significant positive correlation between a neurotic personality (as measured by the score for openness in the Big Five factor model) and the percentage of words in the anxiety category.

(1e) There will be a significant prediction of the sentiment score (sentiment scores for emojis range from -1 to +1, with 0 being neutral (Novak et al., 2015b)) by extraversion (as measured by the score for extraversion in the Big Five factor model).

(1f) There will be a significant prediction of the sentiment score (sentiment scores for emojis range from -1 to +1, with 0 being neutral (Novak et al., 2015b)) by neuroticism (as measured by the score for neuroticism in the Big Five factor model).

1.5.2 Gender Effects and Twitter Measures

(2a) There will be a significant prediction of hashtag density (percentage of tweets containing hashtags) by gender.

(2b) There will be a significant prediction of hashtag type (tag vs. commentary – as measured by individual hashtag densities in the hashtag subset) by gender.

(2c) There will be a significant prediction of hashtag type (tag vs. commentary – as measured by individual hashtag densities in the hashtag subset) by language (German vs. English).

(2d) There will be a significant prediction of emoji density (as measured by the percentage of tweets that contain at least one emoji) by gender.

1.5.3 Gender Effects and LIWC Categories

(3a) There will be a significant prediction of positive emotion words (as measured by the percentage of words in the positive emotion word category) by gender.

(3b) There will be a significant prediction of positive feeling words (as measured by the percentage of words in the positive feeling word category) by gender.

(3c) There will be a significant prediction of negative emotion words (as measured by the percentage of words in the negative emotion word category) by gender.

(3d) There will be a significant prediction of swear words (as measured by the percentage of words in the swear word category) by gender.

(3e) There will be a significant prediction of tentative words (as measured by the percentage of words in the tentative word category (see Appendix C, p. 258)) by gender.

(3f) There will be a significant prediction of words related to social concerns (as measured by the percentage of words in the social concerns category) by gender.

(3g) There will be a significant prediction of words related to family (as measured by the percentage of words in the family category) by gender.

(3h) There will be a significant prediction of percentage of words related to friends (as measured by the percentage of words in the friends category) by gender.

(3i) There will be a significant prediction of words related to occupation (as measured by the percentage of words in the occupation word category) by gender.

(3j) There will be a significant prediction of words related to job (as measured by the percentage of words in the job category) by gender.

(3k) There will be a significant prediction of words related to achievements (as measured by the percentage of words in the achievement category) by gender.

(3l) There will be a significant prediction of words related to money (as measured by the percentage of words in the money category) by gender.

(3m) There will be a significant prediction of words related to sports (as measured by the percentage of words in the sports category) by gender.

1.5.4 Gender Effects and Word-Based Measures

(4a) There will not be a significant difference between the lexical diversity of men and women as measured by Carroll's CTTR.

(4b) There will not be a significant difference between the vocabulary richness of men and women as measured by Yule's K.

(4c) German tweets will show a more 'oral-like' style despite Twitter being a hybrid, mostly written, genre (as measured by the percentages of the two conjunctions *weil* and *denn* 'because,' the former being used in a more informal genre and the latter almost exclusively being used in formal language (Wegener, 1999)).

1.6 Organization of the Dissertation

Chapter 1 presents a brief introduction to the topic of the study along with a current state of affairs, highlights the significance to the field, gives important operationalizations of linguistic key terms, and lists all the hypotheses to be tested. *Chapter 2* offers a detailed literature review on the issue of language and gender variation, the perceived stereotypes that are attached to it, computational gender prediction, and language and gender analysis at the word level. Twitter is discussed as a social medium, and past and present research revolving around it, including Germans as Twitter users. Hashtags, their history, and their functions within a tweet are addressed, as well as two different types of hashtags are also introduced. In addition, the origin of emojis is reviewed and tied to sentiment scores. Finally, research on language and personality as well as the Big Five Inventory's history and development, and its modern applications are discussed. *Chapter 3* details the methodology implemented in this dissertation. The hypotheses are listed together with dependent and independent variables to be used in hypothesis testing. Also, data collection, as well as the Twitter-corpus construction, in addition to statistical analysis procedures are discussed. *Chapter 4* includes detailed descriptive statistics on the participants and the tweet corpus, as well as the hand-coded hashtag corpus, and participants' hashtag usage patterns. Chapter 4 also includes all the hypothesis testing and the discussion of the results. Finally, *Chapter 5* includes an overview of the most important findings and contributions, together with a conclusion, implications, limitations of the study, and recommendations for future research.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

An underlying problem for (variationist) sociolinguistic and online/social media research has been one of obtaining reliable demographic information, and metadata in general. Labov (1972) gauged some demographic information by talking directly to his participants and estimating their age in his seminal New York-study.[2] Then the online data boom took off with Internet Relay Chats (IRC), online messengers, message boards and, most recently, social media, such as Facebook, Reddit, and Twitter. Especially the latter presents a problem: there is usually not much demographic information that can be accessed readily without applying natural language processing to obtain it (gender, age, geographic area, social class, etc.), for instance from tweets (Rao et al., 2010). According to Androutsopoulos (2006), this is mainly due to the fact that demographic information from computer mediated communication (CMC) is often unreliable and incomplete and that there is no phonological/phonetic information, which would let us extrapolate gender from it.

However, the strong upside of social media data, and especially Twitter, is that they abound, which is why the United States Library of Congress started a process in April 2010 to curate all tweets from Twitter's inception until now and into the future (Raymond, 2010, April 14). Apart from the sheer volume of data, the observer's paradox (Labov, 1972) is minimized, as publicly available tweets constitute natural language data (in its broadest sense), which is exactly what sociolinguists look for.

[2] While this 'took care of' the observer's paradox, Labov's (1972) approach would probably be considered too crude today (especially guesstimating participants' age) in addition to being in stark contrast to modern research ethics and IRB standards pertaining to, most prominently, informed consent (see for example the American Psychological Association's code of conduct: APA, 2017).

Recently, researchers have employed Twitter to investigate the propagation of dialectal variation (Russ, 2012). Doyle (2014), for example, used a Bayesian method to estimate the conditional distribution of dialect variants given a specific location. He showed that Bayesian inversion is a viable tool to estimate the probability distribution of dialectal variation, such as 'needs done' vs. 'needs doing' vs. 'needs to be done.' His results are reflected in gold-standard corpora such as the *Atlas of North American English* (Labov, Ash, & Boberg, 2008, ANAE) and the *Harvard Dialect Survey* (Vaux & Golder, 2003, HDS). An approach like this, he (Doyle, 2014) claimed, can save money, time, and effort. In the same vein, Eisenstein, O'Conner, Smith, and Xing (2010) applied a Bayesian computational topic model that estimates and identifies words with regional affinity, linguistic regions (Boston, Northern California, New York, Los Angeles, and Lake Erie), and the relationship between regional variation and topic variation that ensues. They thus provided a first step for modelling linguistic variation under an unsupervised methodology paradigm with raw text.

Similarly, Eisenstein, O'Connor, Smith, and Xing (2014) illustrated the diffusion of lexical change, which is driven by social media and the ensuing informality of conversations. They analyzed a dataset of 107 million tweets (from 2.7 million distinct users) to get to the bottom of how emoticons, abbreviations, phonetic spellings, and other neologism are propagated online (Eisenstein et al., 2014). Their findings show that although computer mediated communication on Twitter makes it easy to overcome great distances, it is a good mirror of the existing linguistic regions of spoken American English (cf. Johnstone, 2010; Kurath, 1949; Labov et al., 2008; Tagliamonte, 2006b) because "the adoption of new written forms is often sharply delineated by geography and demographics" (Eisenstein et al., 2014, p. 1). They also show that cities with similar racial make-up are more prone to share linguistic influence, as racial demographics play a more pronounced role compared to geography when it comes to the diffusion of lexical change (Labov, 1972).

In a related strand of research, Eisenstein (2013) investigated why ‘bad language’ (i.e. misspellings, abbreviations, etc.) is so prevalent in Twitter data (and how lexical diversity online compares in relationship to other established corpora (see also Dresner & Herring, 2010; Herring, 2008, 2010; Herring & Androutsopoulos, 2015). There have also been attempts to realize morphological studies on Twitter and clippings in American English (Reddy & Stanford, 2015).

Bamman, Eisenstein, and Schnoebelen (2014) looked at gender identity and lexical variation on Twitter, gender has long been an important topic in sociolinguistics, especially pertaining to differences in the use of language between males and females (cf. Coates, 1997; Eckert & McConnell-Ginet, 1992; Holmes, 2001, 2006; Lakoff, 1975; Maltz & Borker, 1982). Before computational linguistics gained any real momentum, sociolinguists routinely looked at corpora with latent variables (e.g. socioeconomic status) in mind. Thanks to new computational methods (e.g. Rao et al., 2010), the traditional approach can be reversed and predictions about gender can now be made with raw text input (Bamman et al., 2014).

2.2 Language and Gender

Long standing research on language and gender has suggested that the earliest paradigms built on a clear-cut, dichotomous explanation of how males and females use language differently was flawed at best, e.g. the female deficit approach (Lakoff, 1975); see also the seminal works of Coates (1997), Eckert and McConnell-Ginet (1992), Eckert (2011a), Goodwin (1988), Holmes (1984), Maltz and Borker (1982), and Tannen (1990a, 1991).

While we know that gender, as a social construct, is enacted, i.e. one that allows men and women alike to adjust to various social situations rather than being a fixed state, a quantitative analysis of German Twitter data along with demographic information will shed new light on just how many of the previous stereotypes hold true (or don't) and what new patterns of interactions between language and sociological variables emerge (especially also when factoring in emotion words, and emojis). As extensive, previous research has shown, gender stereotypes do not always hold true and/or are highly situational (e.g. Eckert, 1989; Holmes, 1995). Our new understanding of the gender dichotomy and language notwithstanding, if we want to use gender for new ways of analyzing language, i.e. online data, we have to rely on gender-specific features to get the best results. Myriad studies have used a number of select linguistic features to automatically detect gender in online data with tremendous success: As early as (2007), Argamon et al. built a predictive model of gender, which was 80.5% accurate.

2.2.1 Gender Prediction and Computational Linguistics

As the relationship between language and gender is highly situational (Eckert, 2011a, 2011b; Holmes, 1984, 2001, 2006; Lakoff, 1975), Hastie, Tibshirani, and Friedman (2009) clustered authors who used similar sets of words together. By clustering users who were linguistically similar, they were able to confirm previously known/assumed features of male vs. female use of language. The clusters were formed with the Expectation Maximization algorithm, which groups users by similarities in their word usage. Some stereotypes about male and female use of words were confirmed: taboo words, for example, are used more by male dominated clusters (T. Hastie, Tibshirani, & Friedman, 2009).

Cheng, Chandramouli, and Subbalakshmi (2011) predicted gender with an accuracy of 85.1%, using a feature space consisting of 545 psycho-linguistic, and gender-preferential cues in conjunction with stylometric features. Among these features were 68 categories from LIWC, along with statistical metrics such as Yule's K. They used the Reuters newsgroup dataset, as well as the Enron email dataset to build and test their prediction algorithm. Peersman, Daelmans, and van Vaerenbergh (2011) predicted gender and age in chat text from NetLog with an accuracy of 88.8% based only on tokens and character-based features. Filippova (2012) achieved prediction accuracy of gender of YouTube users on the basis of comments of around 90%, depending on participants' age. Burger et al. (2011), who linked 184,000 Twitter accounts to blog profiles with demographic information, reached 76% accuracy (based on a single tweet!!), which, compared to human prediction performance (68.7%), shows that automatic prediction of gender is more accurate than human judgment. Miller, Dickinson, and Hu (2012) landed an unprecedented accuracy rating of 99.3% using *N*-gram[3] features in tweets (if the tweet had at least 75 characters) — even at 25 characters, the prediction accuracy was still 97.6%. Deitrick et al. (2012) were not too far off either scoring a 98.5% accuracy, using only a limited dataset of 1,484 tweets and a list of 53 *N*-gram features. Fink, Kopecky, and Morawski (2012) achieved 80% accuracy using only unigrams, LIWC categories, and hashtags with Support Vector Machines.[4] Finally, Kokkos and Tzouramanis (2014) achieved gender-prediction accuracies of 92.2% for Twitter and 98.4% for LinkedIn data also using SVMs. Like Cheng et al. (2011), they included function words and content-based features, i.e. words related to positive and negative emotional states from LIWC, to train their classifier. Naturally, no model is perfect, which is why 100%

[3] According to Jurafsky and Martin (2009), "an *N*-gram is an *N*-token sequence of words" (p. 85). Thus, a unigram is a single word, and a bigram is a sequence of two words, such as "I will," or "will go," for example.
[4] Support Vector Machines (SVMs) are supervised machine learning models often used for text classification (Tong & Koller, 2002).

accuracy is hard to achieve. There will always be a small percentage of unexplained variance, which is not captured by the predictors included in the model. Bamman et al. (2014) use the 10,000 most frequent lexical items as independent variables (predictor) and gender as the dependent categorical variable (outcome). The traditional approach of using gender as an independent variable in a logistic regression model was therefore reversed. Gender is not known and can therefore be predicted using text (lexical items) as predictors. The authors achieved 88.0% accuracy, which, compared to previous work (e.g. Burger et al., 2011) was an improvement by almost 8%. In a second step, they used a statistical hypothesis test to determine the fraction of men and women who use a specific lexical item compared to all people who use it. Even then, they found that more than 500 items are significantly associated with gender (Bamman et al., 2014). In addition, Bamman et al. (2014) cluster authors together to get a picture of which authors use similar sets of words (they use a probabilistic cluster algorithm based on the Expectation Maximization framework (T. Hastie et al., 2009)). They found that gender is correlated both with linguistic resources and the composition of the social network (cluster). Since they used probabilistic clustering and not hierarchical, or k-means clustering, they were able to show that over 1,200 men were part of female dominated clusters and over 1,000 women were part of male-dominated clusters. They do this in keeping with previous and long-standing empirical research and the assumption that situated meanings characterize gender and language (Crawford, 1995; Eckert, 2011a; Mulac et al., 2001).

Taking into consideration the levels of accuracy when inferring gender from online data, it makes sense to use similar linguistic features (LIWC categories) in this study to investigate how gender patterns in respect to language, given the success and validity they had in the automatic detection of gender in previous studies. Since, the above studies looked at gender-discriminating linguistic features at the word level, these features are important here as well. However, the goal here is not to predict gender automatically by using the same categories used in the studies mentioned above as predictors in a binary classification algorithm, but rather to take a reverse approach to see whether gender is a significant predictor of classification features, such as Yule's K, or specific LIWC categories included in the hypotheses.

2.2.2 Language and Gender Analysis at the Word Level

Many studies have successfully analyzed language, including male/female differences, not only at a broader contextual sentence level, but also at the phrase and word level. While a word-based approach can be fraught with problems due to the lack of context and the inability to recognize sarcasm for example (a feat that, even today, can only be accomplished automatically through visual semantics and neural networks (Schifanella, Juan, Tetreault, & Cao, 2016)), a solid strand of research now suggests that we are, in fact, able to glean insights about individuals' underlying emotions, thoughts, and motives and their language use in great detail by using a categorized word-counting approach online and elsewhere (see for example the works of Correa, Hinsley, & de Zúñiga, 2010; Golbeck, Robles, Edmondson, et al., 2011; Gottschalk, Stein, & Shapiro, 1997; Iacobelli, Gill, Nowson, & Oberlander, 2011; Mairesse et al., 2007; Misersky et al., 2014; Pennebaker, Booth, et al., 2007; Pennebaker & Francis, 1996; Pennebaker & King, 1999; Pennebaker et al., 2003; Schwartz et al., 2013; Tausczik & Pennebaker, 2010; M. Wolf et al., 2008; Yarkoni, 2010) and that men and women do "adopt different and almost unique gender-based behavioral patterns in communication" (Kokkos & Tzouramanis, 2014, p. 3). Studies looking into specific words/word categories were able to confirm some of the stereotypes about female language use at the phrase-level such as tentativeness (Newman et al., 2008).

In English, women used more intensive adverbs (e.g *very*, *extremely*), more conjunctions (e.g. *but*), and more modal auxiliaries (*could*, *may*, *might*), which could indicate a question mark in the statement (Biber, Conrad, & Reppen, 1998; McMillan, Clifton, McGrath, & Gale, 1977; Mehl & Pennebaker, 2003; Mulac et al., 2001). Male speech, on the other hand, has been associated with a higher frequency of swear words, longer words (> 6 letters), higher use of articles, and overall more references to location (Mehl & Pennebaker, 2003; Mulac, Lundell, & Bradac, 1986). Koppel, Argamon, and Shimoni (2003) discriminately separated male and female authors in a sample from the British National Corpus (BNC), encompassing fiction and non-fiction; their prediction algorithm achieved roughly 80% accuracy. In a similar vein, Biber et al. (1998) used parts of speech to investigate if a given text sample was more involved (more pronouns, present-tense verbs) or more informative (more nouns, long(er) words). They found that the language of females was more involved compared to the language of males (Newman et al., 2008). Newman et al. (2008) directed their focus to gender differences in language use and the word categories included in the English LIWC dictionary.[5] Their findings showed "small but consistent gender differences in language use" (Newman et al., 2008, p. 229). The women in their study used language more to discuss people, what they were doing, and to communicate internal processes to others (including doubts). Further, the list of words that women use more than men also comprised thoughts, emotions, senses, negations, as well as present and past tense verbs (Newman et al., 2008). Men, on the other hand, used language predominantly to label external events, objects, and processes. They also discussed occupation, money, and sports more

[5] For a detailed introduction to Linguistic Inquiry and Word Count (LIWC), see Chapter 3, Section 3.3.1.

than women. On the word level, they used numbers, articles, prepositions, and long words more in addition to more swear words. Interestingly, Newman et al. (2008) did not find discriminate differences in male and female use of references to sexuality, anger, time, use of first-person plural, overall number of words, and qualifiers (exclusion words such as *but*, *although*) looking at an archive of electronic text samples from 70 studies from 22 laboratories in the United States, New Zealand, and England. As mentioned above, further studies in cognitive psychology, computational linguistics, and computer forensics show that women and men do adopt different, almost unique gender-based behavioral patterns in communication (Kokkos & Tzouramanis, 2014). Kokkos and Tzouramanis (2014) used content-based features and traditional linguistic features to analyze potential gender differences and found that men showed distinct patterns of more marked expressions of independence and hierarchical power while women used more emotional language, intensive adverbs, and affective adjectives (*quite*, *adorable*, *charming*, *lovely*).

Finally, Schwartz et al. (2013) used Facebook data to investigate topic and word use based on sociological variables, such as age, and gender. They were able to corroborate many previous findings on gender-related word use (also using LIWC): Females used more emotion words in general (e.g. 'excited') and first-person singular pronouns ('I') as well as making more references to psychological and social processes (e.g. 'love you'). Males, on the other hand, used more swear words and made more object references (e.g. 'xbox'). In terms of age, the youngest participants used significantly more slang, emoticons, and Internet speak (e.g. 'idk,' 'lol'). Topics also progressed with age (school-related topics for 13-18-year olds, college-related topics for 19-22-year olds) (Schwartz et al., 2013). However, the emotion-word category has produced contentious findings in research. While Mulac, Studley, and Blau (1990) as well as Thomson and

Muracher (2001) showed that women did use more emotion words, Mulac, Seibold, and Farris (2000) showed the opposite investigating male and female managers in a work environment. In light of these conflicting findings, it seems important to sort out what kind of emotion words were used: Mehl and Pennebaker (2003) suggested that women do use more positive emotion words, while men used more words related to anger distinguishing the kinds of emotion words in question.

2.2.3 Gender Stereotypes

What makes the studies above so interesting is that they achieved accurate results, without factoring in context and pragmatics, two important aspects that have guided gender-related, linguistic research for decades. Linguists have tried to debunk a clear-cut gender dichotomy in terms of language use, but, in these studies, the outcome variable was exactly that: dichotomous. The rationale for the above-mentioned researchers was that they wanted to predict gender with as much accuracy as possible. To do that, they needed to find categories and linguistic features that discriminately divided the two genders into two classes. Previous research has focused much of its attention on language features that are easily associated with gender stereotypes and the broader context in which such language occurs, e.g. males have been shown to use more words related to sports (Newman et al., 2008), while women have been shown to use more positive emotion words (Schwartz et al., 2013). Conversely, language categories, which are less obviously related to gender issues, such as pronouns, and other function words, have gone unnoticed (Newman et al., 2008; Pennebaker & Stone, 2003).

It needs to be mentioned that the dichotomies found in these newer studies are not 100% absolute and include an error term, which is most likely the human being itself. This means that humans vary in their language use, which is why their natural use of language should not be mistaken for a desire to be politically correct or to conform to a specific genders' language use. They just focus on a wider angle including many linguistic variables to make their prediction algorithms as accurate as possible. While massive big data analysis supports clear-cut gender categories in certain contexts, here, the goal is certainly not to perpetuate or refute stereotypes of any kind but rather to quench the thirst to understand how language patterns in respect to gender, and other demographic variables, such as age, irrespective of any stereotypes. The questions raised in this study seek to answer *how* males and females use language differently on German Twitter not *why* they do it. Gender differences reflect a complex, multi-causal interplay of social, psychological, and situational factors, some of which will also be addressed here. The label, *gender*, is thus attached to the socially enacted gender roles, male or female, not the biologically assigned sex (Crawford, 1995). Let us recall at this point that gender prediction algorithms predict socially constructed and enacted gender not biological sex. Further, identifiable, gender-specific features are not sought to serve to divide the two genders into groups, rather to be able to understand what objective linguistic, and personality features differentiate them. This is where things become contentious and counterintuitive to some extent. "[G]ender dichotomized and decontextualized" as Crawford (1995) put it. I partly agree with her on the notion of gender being a social construct and, as such, not devoid of meaning, but I want to call attention to the fact that there is now enough evidence from studies that bridge the gap between research that looked into language variation pertaining to gender, and new, more recent research, in which gender is, in fact, quantified and operationalized as a dichotomous variable, and thus made predictable with good accuracy across genres and media in certain contexts, as we have seen above. Again, the results from these newer studies are proof of the advancements in computational linguistics, but do not negate the findings from previous research, as they simply present a different angle. Naturally, we should not forget that gender, as a global category, depends on other social variables and context (provided by demographic information for instance) such as educated/uneducated, articulate/inarticulate (vocabulary richness), formal/relaxed (Bamman et al., 2014; Eckert, 2008) — just think of the complex interactions of the gender variable with local categories of school-oriented 'jocks' and the more anti-social 'burnouts' (Eckert & McConnell-Ginet, 1995).

2.2.4 Gender and Informational vs. Involved Language

Another way to describe gender-based language variation is the use of language that is "informational" vs. language that is "involved" (also referred to as emotional) (Argamon et al., 2003). While "informational" language focuses on propositional content, and the frequency with which word classes, such as prepositions and adjectives, are used, "involved" language is used to create interactions between speakers and their audiences, often with first and second person pronouns, for example (Biber, 1989; Tannen, 1982). Past research has shown a correlation between the "informational" style and males whereas the "involved" style was shown to be correlated to female users in studies on CMC corpora, such as blogs (Argamon et al., 2003; Argamon et al., 2007; Schler, Koppel, Argamon, & Pennebaker, 2006). This distinction has also been described as "formal" or "explicit" vs. "contextual." The former uses word classes such as nouns, adjectives, prepositions, and articles, and is often found more in male language, while the latter uses more pronouns, verbs, adverbs, and interjections, and is often found in female language (Mukherjee & Bing, 2010; Nowson, Oberlander, & Gill, 2005). However, this is again a question of frequency; all English speakers use all word classes. It is safe to say that these language characteristics also stem, at least in part, from gender-construction in Western societies, in which men are 'expected' to be more rational and women more emotional (Lutz, 1986, 1990), which, in turn, shapes men's and women's language (Hall, 1995).

In terms of "topics," a relationship between gender and topical domain has been maintained by sociolinguists since the 1980s (Herring & Paolillo, 2006). While all-female groups seem to engage in "gossip" and talk more about people, relationships, and internal states (Aries & Fern, 1983; Coates, 1989), all-male groups tend to focus more on objects such as cars and computers, and external events (politics and sports) (Coates, 1993; Johnson, 1994). Tannen (1990a, 1991) maintains that women engage more in "rapport" talk, and men more in "report" talk, which can be related to a more "involved" way of talking versus a more "informational" way of talking (cf. Argamon et al., 2003 above; Biber, 1989).

2.3 Twitter

The inception of the internet-age ushered in a tremendous shift in how people communicate. From early versions of CMC such as IRC, instant messengers, and blogs to social media, such as Facebook, Instagram, and Twitter today.

2.3.1 Twitter: Past and Present

Twitter is a micro-blogging platform, which was founded in 2006 and just turned ten in 2016. In the form of 140-character messages (tweets),[6] users disseminate their opinions, emotions, attitudes, and comments online, mostly publicly, and on virtually every topic. Twitter now has ~3,860 employees, over 35 offices around the world (Twitter Inc., 2017f), and turned a first profit (~7 million USD) in Q4, 2015 (Nowak, 2016, January 2). To date, there are roughly

[6] At the time of writing, Twitter began testing an increase of tweet length to 280 characters (Isaac, 2017, September 26)s. This, however, did not affect the data, and resulting analyses in this study.

313 million active monthly users worldwide (Statista, 2016; Twitter Inc., 2017f). As of 2015, 1.3 billion accounts had been created (Wagemakers, 2015, August 3) with up to 44% of Twitter users never having sent a single tweet (Murphy, 2014, April 13). On top of its 313 million active users, Twitter also projects 500 million users who 'just' read Twitter feeds and are not actually signed up, totaling roughly 800 million worldwide users (Soziale Medien, 2016, March 21). Around 80% of Twitter's users access the site on their mobile devices, such as smartphones or tablets (Twitter Inc., 2017f). Within the user base, journalists make up almost a quarter (24.6%) of verified accounts[7] (e.g. journalists, politicians, celebrities, important entities, and companies) (Kamps, 2015, May 25). Further, 83% of world leaders are on Twitter (The Digital Policy Council, 2016, January 23) with 79% of accounts being held outside of the United States, and Twitter now supporting 40+ languages (Twitter Inc., 2017f).

Sending two million tweets/day in 2009, and around 65 million tweets/day in 2010 (Twitter Engineering, 2011, June 30), it took users three years, two months, and one day to generate one billion tweets (Twitter Inc., 2011, March 14). In early 2011, it still took around one week to produce the same amount. Only a couple of months later though, in June 2011, users already tweeted 200 million tweets per day. At the time of writing, people around the world (with tweets from the US leading the way with over 50% (Scheffler, 2014)) are generating roughly 500 million new tweets per day, or ~6,000 per second, (K. Smith, 2016, May 17), cutting the time it takes to generate one billion tweets down to two days. In fact, in 2011, a day's worth of tweets (200 million) could fill a 10 million page book (Twitter Engineering, 2011, June 30). By extrapolation, in 2016/2017, a day's worth of tweets (~500 million) fills a 25-million-

[7] Verified Twitter accounts have a "blue verified badge […] [which] lets people know that an account of public interest is authentic." Twitter verifies accounts by request only upon checking Twitter's verified account guidelines (Twitter Inc., 2017a).

page book, or, for perspective, 20,407.5 copies of Leo Tolstoy's *War and Peace*. Reading this much text would take more than 77 years (Twitter Engineering, 2011, June 30). As of 2016, the tears-of-joy emoji (😂) was the most tweeted emoji, with 14.5 billion tweets (Twitter Data, 2016, March 21),

With roughly a quarter (24%) of US internet users on Twitter (21% of all US adults), tweets are almost equally distributed between men and women (Greenwood, Perrin, & Duggan, 2016, November 2016; A. Smith & Brewer, 2012), are relatively easy to obtain (Eisenstein, 2013), make it convenient to build large datasets, and are publicly available, which makes them practical to use in research, with little to no objections from institutional review boards. As Twitter users skew young (36% of 18-29 year olds, 23% of 30-49 year olds, but still a solid 21% of 50-64 year olds) and to some extent, if only slightly, toward urban areas, 26% urban vs. 24% suburban, vs. 24% rural (Duggan & Brenner, 2013; Greenwood et al., 2016, November 2016), and Twitter being slightly more popular among higher educated users; 29% have college degrees vs. 20% have high school degrees or less (Greenwood et al., 2016, November 2016). Doyle (2014) raised the concern of Twitter data-sets possibly being skewed and noisy, which could potentially make tweets inappropriate for sociolinguistic research. However, as he concedes, citing Labov et al. (2008), "young urbanites tend to drive language change" (p. 8), thus making Twitter a great resource for sociolinguistic research as it provides unedited writing data, which is more reflective of non-standard usage than most corpora.

2.3.2 Germans on Twitter

In 2016, for the first time, Twitter Germany published German user projections amounting to 12 million active monthly users (~3.8% of worldwide total users), not indicating, however, what percentage of those users actively tweets on a regular basis or 'just' follows other users for information (Soziale Medien, 2016, March 21). In addition, the demographics of social media users are slightly different in Germany compared to the United States: true social media laggards, only 50% of German adults with internet and/or smartphone access said that social networking sites were popular. This puts them right behind Japan (51%), and at a tie with Pakistan (50%) at the bottom of a list of 40 surveyed countries (Poushter, 2016, February 22). Germany also has the highest social media age gap among those same 40 countries; 39% of people over 35 use social media versus 81% of people in the 18-34 age bracket (a difference of +42). In the US, 63% of people over 35 use social media compared to 89% of 18-34 year olds (a difference of +26) (Poushter, 2016, February 22). Not surprisingly, Germany only has an active social media penetration of 36% (29 million) overall (still a 4% growth compared to 2015), compared to a whopping 59% (192 million) in the US (up 3% from 2015) (Kemp, 2016). When it comes to education, Germany seems to be the polar opposite of the US: only 42% of highly educated people (those with a university education (Eurostat, 2016)) said they use social media at all and 58% said they avoid them. In contrast, 51% of less educated people (high school or less/only some secondary education (Eurostat, 2016)) use social media on a regular basis (OECD, 2015, 2015, November 15). Compared to the US (Twitter gets 3rd place trailing Facebook, and Facebook Messenger), Twitter is only the 7th most popular social media platform in Germany, trailing WhatsApp, Facebook, and Facebook Messenger among others (Kemp, 2016). These findings notwithstanding, conflicting evidence abounds on the exact distribution of German Twitter users. According to the first big Twitter survey in Germany (Pfeiffer, 2009),[8] which included 2,800 participants, German Twitter users are young (32 years on average),

[8] The results of the actual survey are not available online anymore as the website has become defunct, which is why secondary sources were used (Friedrichs, 2009; Hilker & Raake, 2010).

mostly men (74%), and educated (78% have an academic high school diploma). Women are more prone to have blogs and tweet about private matters, life style, and society, to exchange ideas and to stay in contact with friends, while men mostly tweet about Web 2.0, private matters, and technology, in order to gain and disseminate new information, with four out of five users actively tweeting instead of only following other users (Pfeiffer, 2009 as cited in Friedrichs, 2009; Hilker & Raake, 2010). These results are indicative of the pronounced role terminology plays here: while Eurostat (2016) labels university graduates as highly educated, participants in the Pfeiffer (2009) study needed an academic high school diploma to be considered highly educated.[9]

2.3.3 Twitter: (Early) Research

The very nature of a limited amount of characters (140) has yielded interesting ways to communicate and has, in turn, drawn the attention of a lot of researchers from various fields, including linguistics (Twitter Inc., 2017f). Today, social media are used daily around the world, with Facebook being the most popular social network worldwide, having more than 1.86 billion active monthly users (Facebook, 2017). In comparison, Twitter, which has 'only' 313 million active users (Statista, 2016; Twitter Inc., 2017f), may not be close to Facebook in terms of numbers, but it serves a completely different purpose in a more public space than Facebook: users describe their daily routines, carry out conversations, and share information (Java, Song, Finin, & Tseng, 2007; Naaman, Boase, & Lai, 2010), in a medium, which Naaman et al. (2010) call "social awareness stream," underscoring Twitter's ability to convey quasi face-to-face

[9] This must be considered against the backdrop of Germany's three-tiered school system, in which academic high schools do represent the top-tier and thus the graduates with the highest education (an academic high school diploma is needed to gain access and enroll in universities).

interaction. As Darling, Paul, and Song (2012) have pointed out, Twitter, and most other social media for that matter, do not represent a unified genre, which means that registers can range from inside jokes, and wordplay to official government/political announcements.

After Twitter had been launched, new patterns and structures, such as the extended use of hashtags to structure and make tweets searchable, emerged shortly thereafter (Messina, 2007, August 25). Apart from the 140-character limit, there are two meta-characters, '@' and '#.' The former is used to mention/directly address another user, the latter, the hashtag, which debuted in August 2007 (Messina, 2007, August 25, 2007, October 22), is used to index a tweet, comment on it, or affiliate it with a trending topic (Zappavigna, 2015; Tsur, 2012).

Although Twitter has been claimed to be near real-time conversation (Zappavigna, 2011), tweets must still be hand-typed and written, which makes them a hybrid between spoken and written language. Zappavigna (2015) thus called for further research into the 'turn-taking' system of tweets as it *is* different from spoken language. Honeycutt and Herring (2009) looked into how users converse and collaborate on Twitter. They found that utilizing '@' to address other users greatly influenced and promoted user-to-user interaction and collaboration. Reyes, Rosso, and Veale (2013) looked into how irony is constructed and appropriated on Twitter, partially by using the hashtag "#irony." Pointing out the difficulties associated with irony and sarcasm, which are already challenging features in written text, but even more so in tweets, they used a Naïve Bayes classifier and a decision tree to train models, which then detected irony in any given set of tweets according to four conceptual features, signatures, unexpectedness, style, and emotional scenarios, based on the hashtags, #irony, #education, #humor, and #politics. They reach accuracies ranging from 44.88 to 85.40% (Reyes et al., 2013), which shows that detecting irony and sarcasm has yet to be perfected and standardized.

Corporate apologies are accessible to every user 'following' a given corporation's Twitter account due to the fact that most Twitter accounts are publicly available (Madden et al., 2013). Corporations receive feedback from users through the company's hashtag or, directly, via their Twitter handle (e.g. "@apple"). Page (2014) revealed several strategies employed by companies to save face or to deny any wrong-doing on their part to avoid loss of customers or revenue following bad customer service, for instance. In addition, companies use Twitter extensively for brand communication (Nitins & Burgess, 2014; Stieglitz & Krüger, 2014). A versatile data source, Twitter has also been used to study the sleep-wake rhythm in humans (Scheffler & Kyba, 2016) as well as to make predictions about election outcomes for various (presidential) elections, e.g. in Germany (Tumasjan, Sprenger, Sandner, & Welpe, 2010) or the US (Wang, Can, Kazemzadeh, Bar, & Narayanan, 2012) and in politics as well as activism in general (Maireder & Ausserhofer, 2014). Further, Twitter has been used by professional journalists as an ambient news network (Hermida, 2014). For the social sciences and the humanities, the "computational turn"[10] meant readjustment to new media, bigger data, new computational methods for data analysis, and, in some instances, ultimately also a paradigm shift in terms of prior beliefs and research findings in many social science sub-disciplines due to unprecedented and previously unavailable amounts of data (Burgess & Bruns, 2015). Since social media have become ubiquitous in most peoples' lives, a vast amount of research has been generated (Seargeant & Tagg, 2014; Weller, Bruns, Burgess, Mahrt, & Puschmann, 2014).

The beauty of Twitter-data lies in its accessibility, which is only limited by Twitter's own Terms of Service (Russell, 2013). At the same time, however, tweets are usually not geo-tagged

[10] According to Berry (2012), the concept entails the use of computational methods and technologies in the social sciences and humanities "to shift the critical ground of these disciplines' concepts and theories" (p. 11) and to add to Lazer et al.'s (2009) notion of collecting and analyzing data with an unmatched breadth, depth, and scale (Burgess & Bruns, 2015).

anymore (Twitter, 2014), and user gender has to be obtained by means of automated natural language processing (NLP) algorithms, which predict users' gender based on first names, for example (Rao, Yarowsky, Shreevats, & Gupta, 2010). The same holds true for race, age, and other sociological variables. It is obvious that such demographic information is crucial if we want to make our analyses more generalizable and be able to train more accurate models in order to glean more information from Twitter than just gender, race, and location. These obstacles notwithstanding, research has made tremendous progress, and we are now able to predict those variables with varying degrees of accuracy (Bamman et al., 2014; Burger et al., 2011; Eisenstein, 2013; Eisenstein et al., 2014).

2.3.4 Multilingual Twitter

Twitter has also been researched against the backdrop of the multilingual internet. Researchers tried to model multilingual networks on Twitter to better understand the micro-communities and social networks at work in online social media (Eleta & Golbeck, 2014). In a similar vein, Leppänen, Pitkänen-Huhta, Piirainen-Marsh, Nikula, and Peuronen (2009) investigated how young people in Finland use new media and how their use of computer mediated communication influences language choice and heteroglossia.

As of 2010, the top five Twitter languages were English (50%), Japanese (14%), Portuguese (9%), Malay (6%), and Spanish (4%) (Hong et al., 2011; Semiocast, 2010, February 24). According to Statista (2016), this trend has changed slightly: English (34%), Japanese (16%), Spanish (12%), Malay (8%), and Portuguese (6%). It stands to reason that the same five languages are dominating the Twitter-sphere in 2017 given the fact that there are no recent polls that suggest otherwise. What has not changed is the fact that German does not appear in the top five, and barely makes it into the top ten (Hong et al., 2011; Scheffler, 2014). It is thus time to answer the call for further research into German as a Twitter language (Hong et al., 2011; Scheffler, 2014; Weerkamp et al., 2011).

Working within the framework of systemic functional linguistics, Zappavigna (2011) collected 45,000 tweets following Obama's victory in the 2008 presidential elections to investigate how language on Twitter is used to build community, which is an important concept in sociolinguistics (Eckert, 2012; Labov, 1972). By virtue of Twitter's one-to-many interaction, i.e. nobody is obliged to follow someone back, Twitter feeds are often non-reciprocal. However, it appears that interaction between public and private feeds does happen (Zappavigna, 2011). Utilizing discourse analysis and the linguistic function of hashtags, Zappavigna (2011) was able to show how users affiliate with a current topic on Twitter and how this informs/forms the representation and use of language online (2011).

Scheffler (2014) analyzed a randomized set of roughly 24 million German tweets factoring in geolocation to provide an initial snapshot of German Twitter users' tweet behavior. She found that big metropolitan areas such as Berlin lead the way in terms of overall number of tweets produced. Here data include tweets and replies (20% of tweets in the corpus were replies to previous tweets), with the clear majority of these 'discussions' only being two tweets long. In addition, she considered different registers used on Twitter utilizing conjunctions close in meaning to *because* (*weil, denn, da*) to assess formality. Scheffler's (2014) findings suggest that German twitter messages mirror a more oral-like genre as many users use *weil* more than *denn*, which are more formal and claimed to belong to the written register and being exceptionally rare in spoken German (Wegener, 1999). Scheffler (2014) further claims that since German tweets only make up a relatively small number of tweets overall, collecting a German Twitter sample is "virtually a complete snapshot of German [T]witter messages over a certain time" (p. 2284).

2.3.5 Hashtags

2.3.5.1 Hashtag What art Thou?

The poster child of Twitter, the hashtag has proven to serve a very specific function within tweets, its canonical function being that of 'tagging' to make tweets searchable. Their use was first proposed by Twitter user Chris Messina (@chrismessina) in August 2007. He posted a tweet, "how do you feel about using # (pound) for groups. As in #barcamp [msg]?" (Messina, 2007, August 23) and explained his proposed use in two blog posts (Messina, 2007, August 25, 2007, October 22). A tool originally developed to organize and manage information, hashtags have gained expressive functions that go beyond their originally intended use (Shapp, 2014). The "#" symbol is referred to as "hash" in many computer programming languages, such as Python, and R (Houstin, 2013). It is used to introduce "pseudo-code," comments that explain what a certain section of code does, so someone else reading the code can understand it better, but not part of the actual code that is run by the computer to execute a certain command. It seems as if the #-symbol traditionally marks meta-data, which it also does in hashtags. On Twitter, anything that follows the #-symbol up until the next (white) space becomes a 'hyperlinked' hashtag[11] the user can click on (Shapp, 2014). In previous studies, the focus has often been on how hashtags spread and what mechanisms they use (often through Machine Learning) (Chang, 2010; Huang, Thornton, & Efthimiadis, 2010). Zappavigna (2011) was the first to recognize and stress the pronounced role *searchable talk* plays as something completely new attributable to a "cultural

[11] Etymologically speaking, the "hash" is the symbol # + "tag" to indicate the act of tagging a topic/information to make it searchable and organize it (Shapp, 2014).

shift to interpersonal search" (p. 789). She further contends that hashtags serve as linguistic markers that convey a specific message: "[s]earch for me and affiliate with my value!" (Zappavigna, 2011, p. 789). By creating ambient affiliation, users on social networking sites create sociality in new ways, one in which language still plays a crucial role (Zappavigna, 2011).

2.3.5.2 Hashtag (Linguistic) Functions in a Tweet

Zappavigna (2015) investigated hashtags and their key semiotic function within tweets in a qualitative study. What she found was that (a) hashtags serve as a means to organize, label, and structure 'searchable talk' and (b) that they serve different functions depending on their respective placement within a tweet: the beginning, as a constituent of the syntactic structure (a noun, adjective, etc. preceded by a hashtag), and the end. The interesting thing about hashtags is that they function as a sentence constituent on a syntactic level as well as demographic information on a superordinate level. The novelty and difference to 'traditional' metadata lies in their ability to transcend mere 'aboutness' (Kehoe & Gee, 2011) and serve to form communities (however small) on Twitter (Lin, Margolin, Keegan, Baronchelli, & Lazer, 2013; Yang, Sun, Zhang, & Ei, 2012). Zappavigna (2015) used a 'multifunctional' lens on hashtag use and applied Halliday's (1978) threefold meta-functional framework, experiential, interpersonal, and textual, using the 2013 microblogging corpus HERMES2013, which contains 100 million words. Non-English tweets were filtered out, which yielded 2,699,650 hashtags, 16% of posts containing one, 5% of posts containing more than one hashtag. The concordance software, WordSmith, was used to process the natural language data. To attain descriptive statistics, she collected the ten most frequent hashtags found in the corpus. When using the experiential lens, hashtags serve as a structuring device within communication and indicate the topic/theme of a post (Kehoe & Gee, 2011). As such, hashtags can assume various experiential roles as defined by Halliday and Matthiessen (2004): processes (#watch, #listen, #share), participants (#coffee, #hometown), or circumstances (#ontheroad, #wheninrome). Under the interpersonal lens, hashtags provide an evaluative meta-comment — a use that is very frequent and goes beyond the mere reference of topics as it conveys attitudes and affective stance towards said topics as well (#itishardtotell, #sopumped, #annoyed). Finally, the hashtag functions as punctuation using the textual lens. The # symbol has a particular linguistic marking function. The novelty of this meta-datum is that it works seamlessly within the social media text (Zappavigna, 2015). Depending on their position in the tweet, hashtags can serve as the Theme (information at the beginning of a clause), or New (information contained in the Theme). Zappavigna (2015) concludes that hashtags are very flexible semiotic devices that fulfill various roles in social media text. They are topic markers, convey experience, enact relationships, and organize text.

2.3.5.3 Further Research on Hashtags

In general, research into hashtags has been carried out in several related disciplines. They facilitate "conversation" (Rossi & Magnani, 2012) or "discussion" (Bruns, 2012; Bruns & Burgess, 2011). Some quantitative work on hashtags focuses on how hashtags are propagated and spread within social networks (Bastos, Galdini Raimundo, & Travitzki, 2012; Cunha et al., 2011; Ma, Sun, & Cong, 2012; Romero, Meeder, & Kleinberg, 2011; Tsur & Rappoport, 2012). Cunha et al. (2011) found the shorter the hashtag, the more likely and successful its propagation, for instance. As part of the discourse, they are added demographic information, which the users get for free as it were. They are special in that they are embedded into the clausal structure of the tweet (Zappavigna, 2015). Many questions involving hashtags still await an answer though: e.g.

do the same patterns hold true for other languages and are the semiotic/linguistic functions the same for the 'spots' within a post that Zappavigna mentions (2015). Through their omnipresence, metadata nature, and searchability, hashtags serve another superordinate function: self-branding and micro-celebrity. Page (2012) looked at a dataset of 92,000 tweets and found that "self-branding and micro-celebrity operate on a continuum, which reflects and reinforces social and economic hierarchies which exist in offline contexts (p. 181)." Through hashtags, users can direct the attention of other users to their products (corporations) or personal statements, questions, and opinions (Page, 2012). Zappavigna (2011) established the notion of "ambient affiliation," which conveys hashtags' attributive function as linguistic markers. Their function revolves around the idea of organizing and connecting group discussions to topics and affiliates readers and writers of tagged tweets to an ambient community of other Twitter users tweeting about the same topic (Shapp, 2014; Zappavigna, 2011). Zappavigna's (2011) concept of ambient affiliation nicely connects to Shapp's (2014) distinction of tag and commentary hashtags, with the former connecting to an ambient discussion, while the latter "rather function on a level local to the tweet" (Shapp, 2014, p.12).

2.3.5.4 Tag Hashtags

Tag hashtags are canonically used to tag topics, which is done by naming a concrete entity such as a person (#Obama, #Hillary), a place (#Paris, #TheHaven), a company (#Apple, #AMC), or an event (#Thanksgiving, #SummerOlympics). Often used for publicity, these hashtags are intentionally used by companies for advertising purposes. While these tags can be directed at different levels of the public ranging from tags relevant to tweets for the general public all over the world, or country, they also tag entities that are only relevant to certain individuals or a select group of Twitter users/followers. For example, #Paris is a hashtag that speaks to a large audience while #BSU has a much more focused local use and addresses students, faculty, alumni, and staff at Ball State University. Via the tagging function, hashtags organize tweets about a topic, for example to connect one's own tweets to other Twitter users' tweets about the same topic, for example: #Imwithher, #publicprivacy, #blacklivesmatter. Tag-hashtags are also used to organize your own tweets or those of your follower-network, for example: #familyreunion17, #classof2018, #anniversary (Shapp, 2014). Cunha et al.'s (2011) finding ties in nicely here: tag hashtags often seem to be fairly short; one word (#obama), sometimes three (#blacklivesmatter). This makes their success and propagation much more likely and affiliates them with an ambient community.

2.3.5.5 Comment(ary) Hashtags

While the hashtag's main function has remained one of organizing tweets about certain topics, another prevalent hashtag-function has been on the rise for a while: one that Shapp (2014) has aptly dubbed the commentary hashtag. As their name suggests, commentary hashtags serve to "add additional meaning to the main semantic content of the tweet, and are not intended to practically connect the tweet to others that use the same hashtag" (Shapp, 2014, p. 7). Usually, commentary hashtags add an evaluation to what the author of the tweet just said. Routinely, this process adopts the following syntactic pattern: "Text body of the tweet in a sentence. #evaluation." For example, "When someone tells u its not safe to travel to a foreign place alone just cause ur female. #ridiculous #wearenotincapable." Here, #ridiculous qualifies the semantic content of the tweet and elucidates how the author feels about the content rather than trying to connect it with other tweets. Interestingly, the second hashtag, #wearenotincapable, syntactically, is a full sentence in the form of hashtag. The main difference to ‘regular’ commentary hashtags [read: hashtags that evaluate the main body of the tweet] is that these, often longer, hashtags add additional information or an additional idea, which could stand alone separate from the main body of the tweet. In contrast, the main body of tweets with evaluative commentary hashtags can usually stand alone (Shapp, 2014). Since many of these hashtags are very specific and/or syntactically complex it is unlikely that they are used to organize and connect hashtags with similar topics. This is interesting since some tweets do not contain an actual tweet body (text) and the hashtag(s) fulfil the function of carrying the semantic content of the message.

2.3.5.6 Hashtags and Gender.

Herring and Paolillo’s (2006) study on “informational” vs. “involved” language on weblogs showed that personal journal or “diary” blogs contained more female stylistic devices, while “filter” type blogs (knowledge management) contained more male stylistic features. According to Shapp (2014), “tag and commentary hashtags map very closely to the distinction made between knowledge management and diary blogs” (p. 14). It follows that tag-hashtags are used to organize data, whereas commentary-hashtags are used more for self-expression and to be interactive. Shapp (2014) also found that gender is, indeed, being performed in hashtags with males using more tag-hashtags (“informational,” “filter”) and females using more commentary-hashtags (“involved,” “interactional”).

The goal here is not to investigate users’ motivations for the use of either or both kinds of hashtags. Rather, the two-fold distinction is crucial for hand-checking and coding hashtags in the participants’ samples as either tag or commentary hashtags. That way, hashtag densities for tag-hashtags and commentary-hashtags can be calculated and tied to gender to see if Shapp’s (2014) and Herring and Paolillo's (2006) findings can be confirmed in terms of a significant gender split and gender enactment through the type of hashtag that is being preferred by either gender.

2.3.5.7 German Twitter Users and Hashtags.

Germans tend to use hashtags quite a bit, with 18% of German tweets containing hashtags as compared to only 5% in Japanese tweets, for example (Hong et al., 2011). Overall, Germans, on average, seem to be more likely to include URLs and hashtags than users of other languages, which could be indicative of more content-related tweet behavior (Hong et al., 2011). Sometimes Germans even use hashtags to an extent where a tweet only contains hashtags. This could indicate that Germans use them as content substitute and/or with a different function in mind than Twitter users of other languages — one that diverges from the original intended use of the hashtag, that of indexing and "categorizing tweets by keyword" (Twitter Inc., 2017g).[12] Furthermore, a German Twitter user writing a hashtag in English is expected to do so intentionally to contribute to a larger audience and, accordingly, create ambient affiliation (Zappavigna, 2011). Following the paradigm of ambient affiliation, it seems reasonable to assume that an English hashtag in a German tweet is more likely to be used as a tag-hashtag rather than a commentary hashtag.

2.3.6 Emojis

Emojis are essentially emoticons 2.0. Emoticons, which are a short sequence of characters made up of punctuation symbols such as :), :(, or :-), and were potentially first used in the 17th century by a Slovak notary to express satisfaction with the financial affairs of his town

[12] As a best practice, Twitter recommends using maximally two hashtags per tweet (Twitter Inc., 2017g).

(Votruba, n.d.). In the 19th century, an emoticon might have been used in a speech by Abraham Lincoln (J. Lee, 2009). In the digital era, their first use can be traced back to Scott Fahlman, a professor at Carnegie Mellon University, who used them on a computer science bulletin board system on September 19, 1982, to distinguish between jokes and serious messages (Fahlmann, 1982; Krohn, 2004). After that, emoticons spread like wildfire and were extended by hugs, kisses, and more with characters found on standard computer keyboards (Hogenboom et al., 2015; Novak et al., 2015b).

2.3.6.1 Emoji Origins

Having first appeared in Japan to simplify digital communication, “emoji” literally means “picture character.” Its form can be derived by compounding the Japanese words, *e* ‘picture’ and *moji* ‘written character’ (Evans, 2017). After several implementations, Apple™’s support for Emojis on the iPhone™ in 2010, helped them skyrocket in popularity across the globe (Novak et al., 2015b). As of November 2017, there were 2,623 individual emoji characters (encoded in Unicode 10.0) (Unicode.org, 2017). Emojis have come a long way from being used almost exclusively by teenage girls to being used by virtually everyone (Hutchins, 2015, October 14). The Emogi Research Team (2015, September 14) found that nearly 92% of the US population use emojis on a regular basis. While age does not play a major discriminating role, people over 35 still identify as frequent users, and people under the age of 25 use emojis just as much as people between the ages of 25 and 29 (Hutchins, 2015, October 14), gender, in contrast, does play a role, with women using emojis more than men on average (Hutchins, 2015, October 14; SwiftKey, 2015, April 21b). The 2016 emoji report, in which the Emogi Research Team compiles current emoji usage patterns, consolidated previous findings adding some more details, such as that there is an almost equal gender split for medium users (send messages daily, 49% of which contain emojis). However, among light users (send messages a couple of times a year up to weekly, 37% of which contain emojis) males dominate slightly (55%). Conversely, females dominate the heavy user group with 56% (send several messages per day, 56% of which contain emojis). All three groups would like a higher emoji variety beyond the standard set (Emogi Research Team, 2016, November 16). In Germany, emoji frequencies are comparable to most other developed countries (USA, France, Spain, Italy) only differing slightly in the kinds of emojis that are used (SwiftKey, 2015, April 21b).

2.3.6.2 Emoji-Related Research

Not surprisingly, emojis have received as much attention from researchers as their more simplistic predecessors, the emoticons. Wolf (2000) investigated gender differences in emoticon use. Krohn (2004) looked into how different generations use emoticons, claiming that the younger generation [read: millennials] use emoticons ubiquitously even in academia. He suggests that the reason young people do not stick to traditional (email) guidelines is because "[t]oday's college students have never known a world without computers. For them to communicate electronically is natural" (Krohn, 2004, p. 5), Derks, Bos, and von Grumbkow (2007) investigated the importance of social context and the use of emoticons finding that people tend to use a lot more emoticons in socio-emotional contexts compared to fewer emoticons in task-oriented settings. They attribute that, in part, to societal norms, according to which it is more acceptable to show emotions towards friends than towards colleagues (Derks et al., 2007). This is mirrored in Wall, Kaye, and Malone's (2016) findings, whose survey participants reported that emoticons/emojis are inappropriate in professional contexts, using more emojis in texts and on social media than in emails. Schnoebelen (2012) did research on stylistic variation using emoticons on Twitter, showing that there are distinctive differences depending if the emoticon uses a ‘nose’ as in ;) vs. ;-), and finding that non-nose emoticons are leaning towards less standard language (e.g. more taboo words, non-standard spellings), while emoticons with noses are leaning towards more standard language use (Schnoebelen, 2012). Dresner and Herring (2010) looked at emoticons’ illocutionary force. Even before emoticons fully grew into their new appearance as emojis, researchers had already investigated if and how emoticons are used in different cultures by clustering countries with multidimensional scaling to reveal that emoticons like :) (horizontal style) are more common in the US and Europe, whereas emoticons such as ^_^/^^ (vertical style) are much more common in Asian countries (Park, Barash, Fink, & Cha, 2013). The horizontal vs. vertical style distinction has arguably become obsolete with the introduction of emojis. Since both emoticons and emojis perform the same important paralinguistic functions (Dresner & Herring, 2010; Kelly & Watts, 2015), Pavalanathan and Eisenstein (2016) attribute emojis’ replacement of emoticons to their overall success in this paralinguistic role. They (Pavalanathan & Eisenstein, 2016) found that, in their Twitter data, the frequency of emoticons decreased drastically, while the usage of emojis increased tremendously, during the same time span. Thus, I venture that the findings pertaining to emoticons may be adapted to emojis.

Stark and Crawford (2015) looked into how emojis are used in the workplace, Fullwood, Quinn, Chen-Wilson, Chadwick, and Reynolds (2015) considered how text speak (including emojis) relates to personality perception under a psychological lens, and Hudson et al. (2015) examined how gender and emoji use is related to jealousy on Facebook. Emojis’ undeniable ubiquity seems be attributable to their ability to express feelings more accurately than words;

84% of female and 75% of male frequent users subscribed to this explanation (Hutchins, 2015, October 14). Churches, Nicholls, Thiessen, Kohler, and Keage (2014) claim the reason behind this is that people react to emojis (especially face-emojis) the same way they react to a human face because of their resemblance to faces, and, thus, the face-sensitive parts of the cortex react in a similar way they would react to human faces. It is easy to see why emojis are used to such a large extent these days: “[The] [e]moji is to text-speak what intonation, facial expression and body language are to spoken interaction.” (Evans, 2015, November 18). Evans (2015, November 18) argues that emojis are sometimes even more powerful than words, as they are able to convey a broad emotional spectrum — one, which would take several words to describe. Depending on their personalities, people also tend to use emojis differently to mirror reality. For example, social and emotional cues are used more often by people who are agreeable (Côté & Moskowitz, 1998) to convey to other people that they are smiling (Kaye, Wall, & Malone, 2016, April 13). Interestingly, peoples’ perception about themselves also plays a role: participants using more emojis were perceived as more agreeable, conscientious, and open to new experiences (Wall et al., 2016). Further, the use of emojis in online-messages has the potential to influence whether one (mostly a male) gets a date or not: since “emojis facilitate a better calibration and expression of our emotions in digital communication […], it is easier for a potential date to gauge your message” (Evans, 2017, p. 34).

Their omnipresence has also made them a welcome tool to gauge the sentiment of a tweet. In this study, emojis are used to establish sentiment scores for tweets and, by extrapolation, for individual users and, by extension, genders.

2.3.6.3 Emojis in This Study

While recent research has shown that smiling or sad face emojis make up 60% of all emojis that are used (SwiftKey, 2015, April 21a), exploratory data analysis revealed that there are now simply too many emojis, which are used in too many different contexts as to be able to confine them to a set of only smiling or sad face emojis for data analysis. For instance, one female participant (**`part_id = 4`**) used the 'black sun with rays' emoji (☀️) the most (it has a sentiment_score of +0.465 and is thus positive), but not any of the face emojis. The standard UNICODE emoji set (Unicode.org, 2017) now comprises 2,623 different emojis. Novak et al. (2015b) included the 751 most used emojis in their study and calculated sentiment rankings for each of them (ranging from negative via neutral to positive on a -1, 0, +1 scale). The emoji ranking and sentiment is based on their (Novak et al., 2015b) findings. We have to keep in mind one caveat: tweets with emojis will always skew positive. Novak et al. (2015b) found a mean sentiment score of +0.365 for tweets with emojis. This does not mean that the sentiment score is not a good indicator of overall sentiment, as it is more precise than a given predetermined list of emojis due to its range and inclusion of a wider variety of emojis. In this study, the sentiment score per tweet is based on the emojis present in the tweet, using Novak et al.'s (2015b) sentiment scores to match the emoji. If there are two or more emojis present in a single tweet, the sentiment score for this tweet is the mean of those sentiment scores.

With the launch of emoji skin tones in 2015, the Unicode consortium introduced five new skin tones based on the Fitzpatrick scale (Davis & Edberg, 2016; Fitzpatrick, 1975) including pale white to dark brown in addition to a *Simpsons* yellow default color (McGill, 2016, May 9). According to McGill (2016, May 9), white people seem to be avoiding the white, or pale white, skin tone for various, mostly political, reasons, while other races readily embraced the new skin tones to represent them. This could also be attributed to the fact that white people feel like they are represented by default yellow anyway (McGill, 2016, May 9). This is important for the context of this study, because I am not going to factor in individual emoji skin tones as another independent variable, and will only look for default yellow emojis. Figure 1 illustrates how the 751 emojis in Novak et al.'s (2015) study are distributed on the sentiment spectrum with no end of the spectrum reaching the extremes (-1, +1).[13]

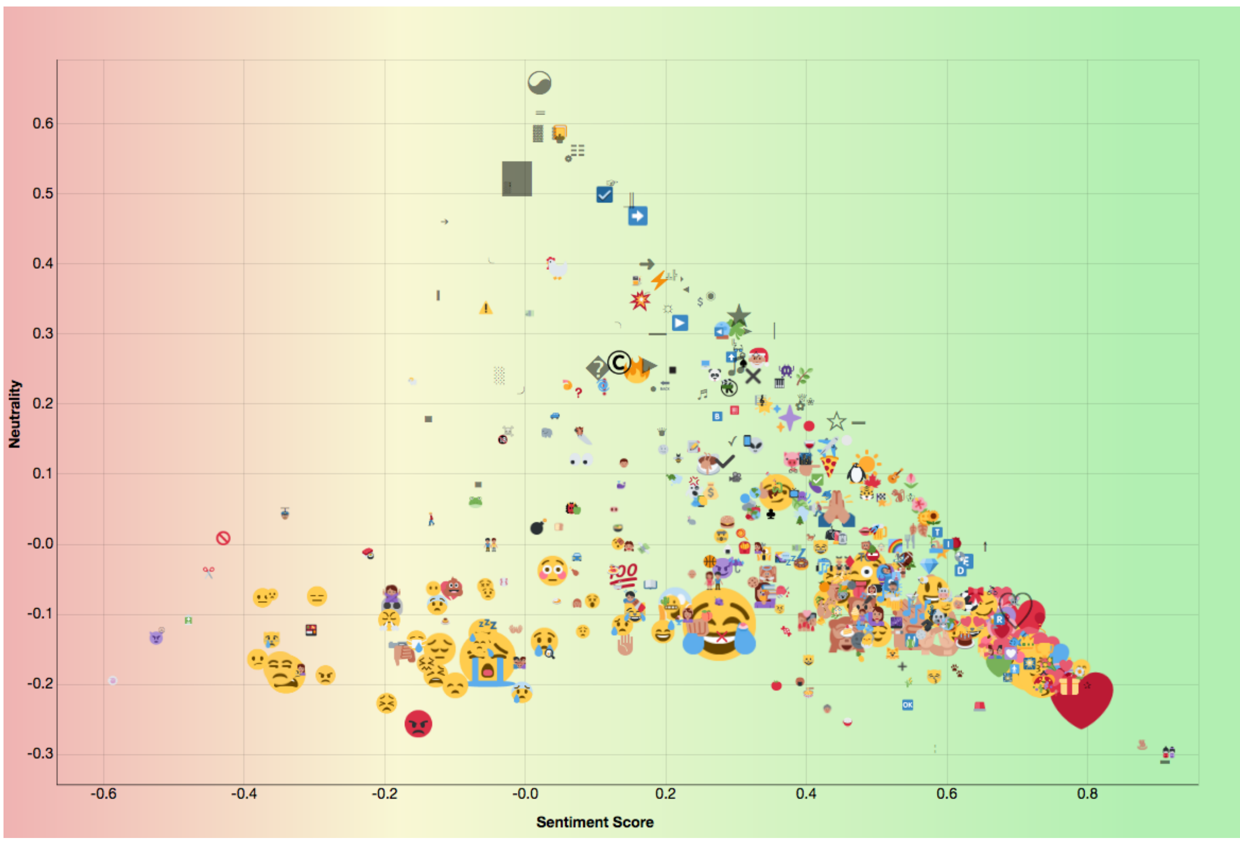

Figure 1. Emoji Sentiment Map (Novak, Smailović, Sluban, & Mozetič, 2016).

[13] For an impressive illustration of real-time emoji use on Twitter, visit Rothenberg's (2013a, 2013b) emojitracker at www.emojitracker.com.

2.4 Language and Personality

Allport and Odbert's (1936) findings that language encodes individual differences in humans established the basis for long-standing research into the relationship of personality traits and language use/linguistic cues. Their lexical hypothesis laid the foundation for the Big Five personality inventory, which has been used many times in research on language and personality and which has become the gold standard in personality research (Mairesse et al., 2007), The Big Five trait taxonomy will be discussed in more detail below. Researchers such as Gottschalk and Gleser (1969) and Weintraub (1989) built on those early studies by investigating how psychological states can be assessed through content analysis and how verbal behavior relates to personality. While assessments of linguistic style were carried out on several levels, including the word level, morphology, syntax, two paradigms were prevalent in psychological text analysis: (1) the psychoanalytic orientation that requires trained raters to assess individual clauses of a sentence (Gottschalk and Gleser, 1969), or Weintraub's (e.g. 1989) approach, in which he compared medical diagnoses with 15 general categories into which he 'inserted' coded words and phrases, or alternatively, (2) a word-based counting system. This paradigm is based on the assumption that "individuals [who] are verbally expressing sadness […] would be more likely to use words such as *sad, cry, loss*, or *alone*" (Pennebaker & King, 1999, p. 1297).

2.4.1 Automated Content Analysis and Linguistic Inquiry and Word Count

With the advent of faster and cheaper personal computers in the mid-nineties, studies, which included automated word recognition, increased in number. Gottschalk, Stein, and Sharpiro (1997) first used automated content analysis of speech for the diagnostic processes in a psychiatric outpatient clinic. While this is only a tangent for the present study, it is worth mentioning as it belongs in the general time line of research focusing on the relationship between content (language) and psychology. Research quickly took off from there. Pennebaker and King (1999) focused on language use as an individual difference and 'revolutionized' the field of language and personality interaction with their software program, Linguistic Inquiry and Word Count (LIWC, which will be introduced in the Methodology, Chapter 3, Section 3.3.1), which was used for the first time in their study (including over 2,000 words coded in 74 different word categories encompassing linguistic dimensions such as function words and grammatical categories, as well as psychological factors such as affective, cognitive, and social processes) and, as will be seen below, has been used in a lot of studies on the subject since (Pennebaker, Booth, et al., 2007). Among their (Pennebaker & King, 1999) most notable findings was a negative correlation between openness and the immediacy LIWC factor (1st person singular, words longer than six letters, present tense, and discrepancies), and a negative correlation between the Making Distinctions LIWC factor (exclusive, tentativity, negations, inclusive) and extraversion as well as a negative correlation with conscientiousness. In terms of individual word categories, neuroticism was positively correlated to negative emotion words and, conversely, negatively correlated to positive emotion words. Extraversion was positively correlated to positive emotion words and total social references. Agreeableness was positively correlated with positive emotion words and negatively correlated with negative emotion words. As for other demographic variables, Pennebaker and King's (1999) most notable finding was that a high score on the immediacy dimension was consistently related to young females with lower SAT scores and exam grades, and whose parents had lower levels of education.

2.4.2 Research Related to Language and Personality

2.4.2.1 Early Research

It is not surprising that a lot of research in the field coincided with the skyrocketing number of online presences, websites, blogs, and budding social media sites (Amichai-Hamburger & Ben-Artzi, 2000; Amichai-Hamburger, Wainapel, & Fox, 2002). It is clear, however, that these studies are over ten years old now, and neither Twitter, nor Facebook had yet been invented. This is necessary to keep in mind as Twitter, as a hybrid genre, might have a different impact on online behavior.

Gill, Oberlander, and Austin (2006), for example, investigated personality in emails at zero-acquaintance.[14] Mairesse, Walker, Mehl, and Moore (2007) claim that, at that point, there had only been two other studies which used automatic recognition of personalities in language data, which is why they focused on linguistic cues for automated recognition of personality in conversation and text. A major shortcoming of their study is the fact that both genres, written and spoken, were obtained in a laboratory setting, which does not aptly represent natural language data. It is those shortcomings that Yarkoni (2010) sought to address in his study on personality and language, in which he extended the analysis beyond the category levels and also investigated the relationship between personality and individual words. To counteract the shortcomings of previous studies, such as written samples from laboratory settings, directed writing tasks, short time spans of data collection, and small sample sizes, Yarkoni (2010) used online blogs for his analysis, which represent a valid written genre and natural language data. Using 66 of the 74 LIWC categories (excluding non-semantic words), his study was in great alignment with previous research: Neuroticism was found to positively correlate with negative emotion words

[14] Zero-acquaintance personality judgment: the rater makes a judgment about a participant [email] with no prior interaction (Cleeton & Knight, 1924).

(e.g. anxiety/fear) and total negative emotions. Agreeableness, on the other hand, showed a positive correlation between positive emotion words and words revolving around social communality (e.g. 1st person plural, family, friends) while at the same time being negatively correlated to negative emotion words and swear words. Diverging from previous research, he (Yarkoni, 2010) found extraversion to be negatively correlated to word categories revolving around goal orientation and work-related achievement (this seems odd and counterintuitive and has, to my knowledge, not been replicated in other studies) and positively correlated to words reflecting social settings or experiences (e.g. bar, restaurant, drinking, dancing). Openness was found to be strongly positively correlated to words associated with intellectual or cultural experiences (e.g. poet, culture, narrative, art). He also found agreeableness to be positively correlated to sexual words (Yarkoni, 2010). His study coincided with one of the first US-German studies on the subject of personality and language (Back et al., 2010).

Küfner, Back, Nestler, and Egloff (2010) looked into the relationship between personality and creative writing (again, convenience sampling as guided data collection procedure in a laboratory setting was used). They had participants (German male/female university students between the ages of 18 and 45) write short stories based on target words along with filling out the Big Five questionnaire. The writing samples were then analyzed with LIWC. Among the first studies that involved social media was Back et al.'s study (2010), which looked into personality and Facebook profiles from US and German social media users. Centering their study around the contrast of the idealized virtual-identity hypothesis versus the extended real-life hypothesis, they found that the users' actual personality is reflected in their online behavior and that they do not feel obliged to create a self-idealized online personality, which is in alignment with the extended real-life hypothesis. Innovative in their approach, they administered the novel short version of the Big Five inventory to German users (BFI-10) and the Ten Item Personality Inventory (TIPI) to US users (Gosling et al., 2003; Rammstedt & John, 2007). Furthermore, they used StudiVZ and SchülerVZ, two German social media sites, as German data source even though Facebook had already been around in Germany. This might make their data susceptible to potential skew, and their reasoning behind not using German Facebook users remains unclear (Back et al., 2010). Continued research has involved more and more social media, such as Golbeck, Robles, and Turner's (2011) study, which bridged the gap between social media and personality research with data from Facebook and taking a more fine-grained approach than Back et al. (2010). Investigations into text messaging as a function of personality traits (Holtgraves, 2011) revealed (and confirmed) significant correlations between several LIWC categories and extraversion (e.g., personal pronouns), agreeableness (e.g., positive emotion words), and neuroticism (e.g., negative emotion words). Interestingly, Holtgraves (2011) also found that linguistic alterations, such as abbreviations, vary according to personality traits and relationship status.

2.4.2.2 Recent Studies on Language and Personality

A more recent study comes from Qui, Lin, Ramsay, and Yang (2012), which, to my knowledge, was the first study using tweets. They use three different sampling methods: snowball sampling, on-campus recruitment, and Amazon's Mechanical Turk. They analyzed a total of 28,978 tweets collected over the period of one month. While they claim that the sample size is comparable to previous studies, they call for extended research, longer sampling time, more participants, and research in a language other than English to discover possible cross-cultural differences. They found that extraversion was significantly correlated with positive emotion words and social process words, while at the same time being negatively correlated to the use of articles, supporting Pennebaker and King's (1999) findings. This seems to point toward extraverts' craving for social attention and a preference for reduced linguistic complexity. They report agreeableness to be negatively correlated to negation words (to be expected and previously found in online blogs, (Nowson, 2006)), swear words, and negative emotion words. Further, Qui et al. (2012) were also able to replicate Yarkoni's (2010) findings (openness is negatively correlated to second-person pronouns, assent words, and positive emotion words in blogs) and Mairesse et al.'s (2007) findings that openness is negatively correlated to past tense verbs in daily language use. Their own, non-replicated findings indicate that extraversion was tied to a higher use of assent words, lower use of function words overall, and fewer impersonal pronouns while openness was positively correlated to the use of prepositions and negatively correlated to the use of adverbs, swear words, and affect words (overall). Participants scoring higher on the neuroticism scale made greater use of negative emotion words and, conversely, used fewer positive emotion words. More agreeable people used fewer exclusive words and fewer sexual words (Qui et al., 2012). The fact that they were able to replicate many of the correlations, positive or negative, indicates that there is a level of consistency between personality factors and language online and offline.

2.5 The Big Five Inventory

2.5.1 Early Development

The Big Five personality inventory was developed after psychologists had tried to come up with an integrative taxonomy on how to measure personality based on and related to language (John & Srivastava, 2001). The Big Five's groundwork was laid by two German psychologists in the earliest attempts to systematically organize the language of personality.

Klages (1926) postulated that a thorough analysis of language can help us understand personality (Digman, 1990). McDougall (1932) had already made the claim that personality may be analyzed along five separate factors: intellect, character, temperament, disposition, and temper. This foreshadowed the end result of half a century of work to conceive a coherent organization for the language of personality (Digman, 1990). Baumgarten (1933), building on Klages' research, investigated personality terms, which are frequently found in German. While their research did not leave a notable mark on German research in psychology, it did influence Allport and Odbert (1936) to embark on an examination of language on their own (Digman, 1990; John, Angleitner, & Ostendorf, 1988), which, in turn, directly impacted research that followed. After that, researchers repeatedly obtained the Big Five traits by applying factor analysis to a number of lists with personality traits (Cattell, 1943, 1948; Digman, 1990; Fiske, 1949; Tupes & Christal, 1957, 1961). Various personality traits percolated gradually into the Big Five inventory as it were (Mairesse et al., 2007). A term probably first used by Goldberg (1981), the Big Five comprise *Gewissenhaftigkeit* 'conscientiousness' (self-discipline, organization, and impulse control), *Verträglichkeit* 'agreeableness' (tendency towards cooperation, social harmony, and consideration of others), *Neurotizismus* 'neuroticism' (tendency to experience negative emotions, anxiety, depression, and anger), *Offenheit* 'openness' (reflects imagination, creativity, and intellectual curiosity), and *Extraversion* 'extraversion' (assertiveness and positive emotionality) (Golbeck, Robles, & Turner, 2011; Goldberg, 1992; John & Srivastava, 2001; Weisberg et al., 2011).

2.5.2 Modern Versions of the Big Five

Several different versions of the Big Five measurement are used today. The most well-known Big Five inventory is the NEO-PI-R (Costa & McCrae, 1992; R. McCrae & Costa, 1990), which comprises 240 items. Its short version, the NEO-FFI (Costa & McCrae, 1992), still comprises 60 items. Yet shorter is the 44-item BFI version of the Big Five measurement (John et al., 1991), which takes around five to ten minutes to complete. As this is still too long for many modern-day research applications, several short versions have been created in an attempt to make the measurement even shorter and economical without sacrificing reliability and validity: the BFI- 25, with 25 items (Benet-Martinez & John, 1998), the BFI-K, which comprises 21 items (Rammstedt & John, 2005), the BFI-S with 15 items (Schupp & Gerlitz, 2008), and the BFI-10 comprising 'only' ten items (Rammstedt & John, 2007; Rammstedt et al., 2012). The TIPI is the US version of the German BFI-10 (Gosling et al., 2003). Although the shortest version, the 10-item BFI-10, only takes one minute or less to complete, it mirrors the five-factor structure very well. It not only has great compliance with the overall BFI scale, but even with the extensive NEO-PI-R (240 items), which makes it an economic instrument to measure the Big Five dimensions reliably and with great validity (Rammstedt et al., 2012).

Today, the Big Five inventory is the gold standard for assessing personality traits (Mairesse et al., 2007); researchers have established the model's general consistency across age, gender, and cultural lines (John, 1990; R. McCrae & Costa, 1990) as well as its validity across different languages (Digman, 1990; John, 1990; R. McCrae & John, 1992; R. R. McCrae, 1989).

2.5.3 Modern Applications of the Big Five

In the wake of the model's general acceptance as the go-to measurement tool for personality assessment, many applications have been tested and found valid: Selfhout et al. (2010) showed that there is a link between personality and who one chooses as a friend on Facebook, with extraversion, agreeableness, and openness all correlating with friendship selection. In other areas of social life, the Big Five have been related to romantic relationships, partner choice, level of attachment, and relationship success (Chamorro-Premuzic, 2007; Shaver & Brennan, 1992). Furthermore, the five factors have been connected to individuals' coping responses, vengefulness, and rumination in the realm of interpersonal conflict (Barrick & Mount, 1993; T. O'Brien & DeLongis, 1996). Beyond social issues and relationships, many studies have tied personality traits, as represented in the Big Five, to a factor that plays a role with peoples' preferences in music (Rawlings & Ciancarelli, 1997; Rentfrow & Gosling, 2003), who people would be more likely to vote for, McCain or Obama (Jost, West, & Gosling, 2009), differences in the personalities of 'dog people' versus 'cat people' (Gosling, Sandy, & Potter, 2010; Perrine & Osbourne, 1998), and prediction of a consumer's preferences for either national or independent brands (Whelan & Davies, 2006). Furthermore, the Big Five have played a valuable role in professional contexts by establishing their usefulness in personality profiles: while Hodgkinson and Ford (2008) report that personality traits influence job performance and satisfaction, Barrick and Mount (1993) were able to correlate specific traits with occupational choices and proficiency (Golbeck, Robles, Edmondson, et al., 2011). In addition, the Big Five have been used successfully to predict entrepreneurial status (Zhao & Seibert, 2006), team performance (Neuman, Wagner, & Christiansen, 1999), and counterproductive behavior (Salgado, 2002).

While extraversion has again and again been found to be a reliable predictor for the use of social media among the Big Five traits (Correa et al., 2010), there is more to the issue than its mere face value would suggest. Researchers showed that there are two competing hypotheses that are linked to extraversion and social media usage: the social compensation hypothesis (introverts gain more from social media due to their personality) and the rich-get-richer hypothesis (extroverts transfer their offline sociability to social media) (Correa et al., 2010; Valkenburg & Peter, 2007). The latter seems to go hand-in-hand with a steady rise of narcissism linked to social media usage, especially among adolescents (Twenge, Konrath, Foster, Campbell, & Bushman, 2008). This development has been confirmed in recent research on social media usage and narcissistic online behavior as a function of personality traits (DeWall, Buffardi, Bonser, & Campbell, 2011; Ong et al., 2011).

Popov, Kosinski, Stillwell, and Kielczewski (2017) built an online tool hosted at the University of Cambridge's psychometrics center, which uses a predictive machine learning model trained on six million social media profiles to predict a user's Big Five personality profile, along with other personality profiles, and demographic information from their social media footprint on Facebook or Twitter.

2.6 Feature Selection

The research in this study and the hypotheses (see introduction Chapter 1 or methodology in Chapter 3) are guided by what has become obvious in prior research mentioned in the literature review and, in particular, in Cheng et al.'s (2011) study as well as elsewhere (Chung & Pennebaker, 2007; Pennebaker et al., 2003): word-based features and function words (among other features) serve as valuable gender discriminators and predictors. Cheng et al. (2011) used

545 features (character-based, word-based, syntactic, structure-based, and function words) and found 157 features to be significant (alpha = .05). Among the word-based features were statistical metrics, such as vocabulary richness (Yule's K), as well as 68 features (word categories) extracted from the LIWC 2007 dictionary. Given their overall good results (> ~85% accuracy), several of their features will be used as measurements here (see Chapter 3). Cheng's (2011) linguistic feature selection overlaps to some extent with Bamman et al.'s (2014), which shows how the same/similar features are used in different analyses. Similarly, Golbeck, Robles, and Turner (2011) found that swear words, social process words, affective process words, positive emotions (among others) are correlate well (some better than others) to the individual Big Five personality traits.

In terms of lexical diversity,[15] previous research has not found significant differences between males and females in Anglo-Australian speakers of English (Alami, Sabbah, & Iranmanesh, 2013). However, it needs to be mentioned that Alami et al. (2013) did not use any specific measures of lexical diversity and investigated an arguably small linguistic sample. They simply define lexical diversity as $LD = \frac{Nlex}{N} \times 100$ thus forgoing more accurate and involved lexical diversity measures. This is why Yule's K (Cheng et al., 2011; Miranda-García & Calle-Martín, 2005; Tanaka-Ishii & Aihara, 2015) and Carroll's CTTR (Carroll, 1964; Hess, Ritchie, & Landry, 1984) are included as metrics. Other word-based features, such as total number of words were also included as well as positive/negative emotion words, swear words, modal verbs and words expressing tentativeness, and words revolving around occupation, social concerns, family and friends, job, achievements, sports, and money since all of those word categories have produced viable results not only in linguistic (Argamon et al., 2007; Bamman et al., 2014;

[15] Lexical diversity measures how vocabulary rich (and thus complex) a text is.

Burger et al., 2011; Cheng et al., 2011; Kokkos & Tzouramanis, 2014; Rao et al., 2010) but also in personality research (Chung & Pennebaker, 2007; Golbeck, Robles, Edmondson, et al., 2011; Golbeck, Robles, & Turner, 2011; Hirsh & Peterson, 2009; Nowson, 2006; Ong et al., 2011; Pennebaker, Booth, et al., 2007; Pennebaker & King, 1999; Pennebaker et al., 2003; Yarkoni, 2010).).

Bamman et al. (2014) found, for example, that while emoticons were said to be used equally by females and males, they are used more by females on Twitter. Another distinctive difference is the inclusion of emojis (instead of emoticons) and their sentiment scores, which is novel and has not been implemented in any of the aforementioned studies (Bamman et al. 2014 used emoticons, for example).

In terms of hashtags, the focus will be on the overall distribution (percentage) based on gender and what type of hashtag (tag or commentary) is being used more by either gender. Herring and Paolillo (2006) warn that gender effects might disappear when other predictors are included in a model[16] — this is what happened in their study on gender differences in weblogs after controlling for the genre of the blog — the risk here is reduced as Twitter does not necessarily represent entirely different genres, rather one unified hybrid-genre in which every user roams. The recently identified "Twitterature" (Aciman & Rensin, 2009) is certainly something to look out for, but given the vast amount of daily tweets it seems safe to assume that the majority of tweets fall within the same non-literary genre and thus data skew is unlikely (Aciman & Rensin, 2009; Rudin, 2011).

---

[16] Simpson's paradox: A significant effect or trend can appear in different groups of participants or data, but can disappear or be reversed when the groups are combined, or a new variable is added to a statistical model (Berman, DalleMule, Greene, & Lucker, 2012, September 25; Wagner, 1982; Wardrop, 1995).

CHAPTER THREE: METHODOLOGY

3.1 Introduction

This chapter first details the description of the research design and the hypotheses to be tested in this study. Then, the data collection and corpus creation will be described, followed by an explanation of the research design, i.e. participant selection, data collection, sample size, and data pre-processing as well as statistical considerations. To reiterate, the goal of this study is to examine interactions between language, gender, and personality of German Twitter users in tweets coming from Germany exclusively.

3.2 Research Hypotheses

As the data-collection approach is twofold (Twitter data in addition to questionnaires with demographic information and personality profiles), the hypotheses fall into four categories: personality and linguistic features, Twitter measures in relationship to gender, gender effects and LIWC categories, and word-based measures as related to gender.

3.2.1 Effects of Personality on LIWC Categories

(1a) There will be a significant positive correlation between an extraverted personality (as measured by the score for extraversion in the Big Five factor model) and the percentage of positive emotion words.

(1b) There will be a significant positive correlation between an agreeable personality (as measured by the score for openness in the Big Five factor model) and the percentage of positive emotion words.

(1c) There will be a significant negative correlation between an agreeable personality (as measured by the score for openness in the Big Five factor model) and the percentage of swear words.

(1d) There will be a significant positive correlation between a neurotic personality (as measured by the score for openness in the Big Five factor model) and the percentage of words in the anxiety category.

(1e) There will be a significant prediction of the sentiment score (sentiment scores for emojis range from -1 to +1, with 0 being neutral (Novak et al., 2015b)) by extraversion (as measured by the score for extraversion in the Big Five factor model).

(1f) There will be a significant prediction of the sentiment score (sentiment scores for emojis range from -1 to +1, with 0 being neutral (Novak et al., 2015b)) by neuroticism (as measured by the score for neuroticism in the Big Five factor model).

3.1.2 Gender Effects and Twitter Measures

(2a) There will be a significant prediction of hashtag density (percentage of tweets containing hashtags) by gender.

(2b) There will be a significant prediction of hashtag type (tag vs. commentary – as measured by individual hashtag densities in the hashtag subset) by gender.

(2c) There will be a significant prediction of hashtag type (tag vs. commentary – as measured by individual hashtag densities in the hashtag subset) by language (German vs. English).

(2d) There will be a significant prediction of emoji density (as measured by the percentage of tweets that contain at least one emoji) by gender.

3.1.3 Gender Effects and LIWC Categories

(3a) There will be a significant prediction of positive emotion words (as measured by the percentage of words in the positive emotion word category) by gender.

(3b) There will be a significant prediction of positive feeling words (as measured by the percentage of words in the positive feeling word category) by gender.

(3c) There will be a significant prediction of negative emotion words (as measured by the percentage of words in the negative emotion word category) by gender.

(3d) There will be a significant prediction of swear words (as measured by the percentage of words in the swear word category) by gender.

(3e) There will be a significant prediction of tentative words (as measured by the percentage of words in the tentative word category (see Appendix C, p. 258)) by gender.

(3f) There will be a significant prediction of words related to social concerns (as measured by the percentage of words in the social concerns category) by gender.

(3g) There will be a significant prediction of words related to family (as measured by the percentage of words in the family category) by gender.

(3h) There will be a significant prediction of percentage of words related to friends (as measured by the percentage of words in the friends category) by gender.

(3i) There will be a significant prediction of words related to occupation (as measured by the percentage of words in the occupation word category) by gender.

(3j) There will be a significant prediction of words related to job (as measured by the percentage of words in the job category) by gender.

(3k) There will be a significant prediction of words related to achievements (as measured by the percentage of words in the achievement category) by gender.

(3l) There will be a significant prediction of words related to money (as measured by the percentage of words in the money category) by gender.

(3m) There will be a significant prediction of words related to sports (as measured by the percentage of words in the sports category) by gender.

3.1.4 Gender Effects and Word-Based Measures

(4a) There will not be a significant difference between the lexical diversity of men and women as measured by Carroll's CTTR.

(4b) There will not be a significant difference between the vocabulary richness of men and women as measured by Yule's K.

(4c) German tweets will show a more 'oral-like' style despite Twitter being a hybrid, mostly written, genre (as measured by the percentages of the two conjunctions *weil* and *denn* 'because,' the former being used in a more informal genre and the latter almost exclusively being used in formal language (Wegener, 1999)).

To address the above hypotheses, different types of data are needed: (1) to investigate linguistic features, participants' tweets are needed as they provide the natural language data, (2) to investigate participants' personality, scores from the Big Five trait inventory are needed to provide measurements on extraversion, openness, agreeableness, conscientiousness, and neuroticism, and (3) participants' demographic information is needed to test the hypotheses revolving around interactions between gender, language and personality.

3.2 Independent (Predictor/Covariates) and Dependent (Outcome) Variables:

3.2.1 Independent Variables

3.2.1.1 Demographic Variables

Gender (dichotomous, categorical). Gender of the tweet authors (male or female), a dichotomous variable from the questionnaire.

Age (continuous). The age of participants was determined through the questionnaire and is measured on a continuous scale.

Zip codes I (numerical). The zip codes of participants' place of birth was determined through the questionnaire.

Relationship status (categorical). Participants' relationship status was measured as a categorical variable (levels = married, single, divorced, widowed)

Education - school (categorical). Participants' education level was measured based on the four common types of schools in Germany (Hautpschule (lowest level), Realschule (mid-level), academic high school (highest level), FOS/BOS (alternative post-vocational training schools)).

Highest level of education (categorical). Participants' highest level of education was assessed with four levels (vocational training, university of applied sciences, university, no degree).

The questionnaire also assessed participants' citizenship and native language, as well as an additional zip code if they had grown up somewhere else that is not their current place of residence. These three variables were not included in the analysis, as all participants were German nationals with German as their first language.

3.2.1.2 Social Media Variables

Twitter usage length in years (categorical). Measures how long participants have been Twitter users.

Twitter update check/day (categorical). Measures how many times participants check updates on Twitter.

Time on Twitter/day (continuous). Measures how much time participants spend on Twitter in minutes.

Hashtag function (categorical). A measure to determine whether the users have an idea about the intended use of hashtags (tagging vs. commenting, or both).

3.2.1.3 Personality: The Big Five Factor Model (BFI-10)

The Big Five factor model comprises *Extraversion* ‘extraversion,’ *Verträglichkeit* ‘agreeableness,’ *Gewissenhaftigkeit* ‘conscientiousness,’ *Neurotizismus* ‘neuroticism,’ and *Offenheit* ‘openness.’ These factors are usually measured with Likert-scales, ranging from one (strongly disagree) to five (strongly agree). Participants’ Likert-scale answers to statements (ten for the BFI-10) for each factor are averaged to obtain individual scores for each factor. Jointly, the resulting five scores represent an individual’s personality (John & Srivastava, 2001). For example, see one participant’s (**`part_id = 1`**) personality profile in Figure 2 below. A high score on extraversion means that an individual is sociable, outgoing, talkative, and assertive, while a high score on agreeableness indicates that an individual is cooperative, helpful, and nurturing. People scoring high on conscientiousness are responsible, organized, hard-working, and reliable. Individuals, who have high scores on the neuroticism trait are anxious, insecure, and sensitive, while people, who score high on openness are curious, intelligent, and imaginative (Golbeck, Robles, & Turner, 2011).

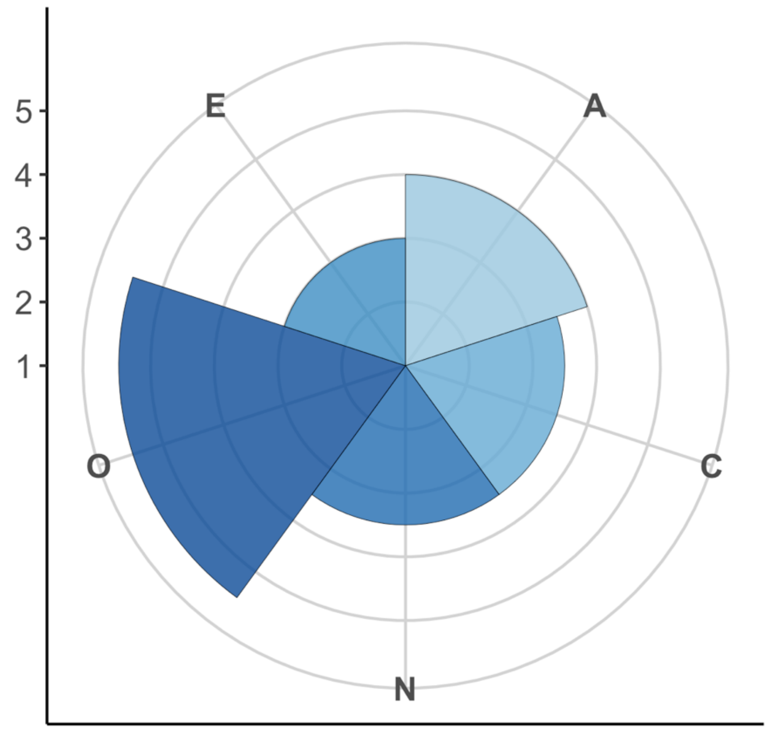

Figure 2. Big Five Personality Profile.

Extraversion (ordinal). Measures participants' level of extraversion on a five-point Likert-scale.

Openness (ordinal). Measures participants' level of openness on a five-point Likert-scale.

Agreeableness (ordinal). Measures participants' level of agreeableness on a five-point Likert-scale.

Conscientiousness (ordinal). Measures participants' level of conscientiousness on a five-point Likert-scale.

Neuroticism (ordinal). Measures participants' level of neuroticism on a five-point Likert-scale.

3.2.2 Dependent Variables

TT_50 (continuous). A measure of how long (in days) it took a user to reach 50 tweets. This is based on Shapp's (2014) span-measure and shows how avid a Twitter user is. A low TT_50 indicates that a user is more avid (more frequently uses Twitter) compared to a user with a high TT_50.

CTTR (continuous). A measure of lexical diversity, Carroll's (1964) corrected type token ratio (CTTR) is based on the traditional TTR[17] with a correction for uneven sample sizes. The higher the CTTR, the greater the lexical variety in a text (Hess et al., 1984); $CTTR = \frac{types}{\sqrt{2\ x\ tokens}}$ where *types* = total number of different words, and *tokens* = total number of words. Since this measure represents a percentage rather than a frequency, there are not going to be problems statistically speaking.

Yule's K (continuous). Yule's K is another measure of lexical diversity (Cheng et al., 2011; Miranda-García & Calle-Martín, 2005; Tanaka-Ishii & Aihara, 2015). The larger Yule's K, the easier the text. $K = 10^4 \times \left[-\frac{1}{N} + \sum_{i=1}^{V} V_i \left(\frac{i}{N}\right)^2\right]$, where N = total number of words, V = number of different words, and V_i = number of different words that occur i times.

[17] The type-token-ratio (TTR) is one of the most well-known measures of vocabulary richness, $TTR = \left(\frac{types}{tokens}\right)$. Since the TTR is susceptible to skew when samples of different sizes are analyzed, Carroll (1964) developed a corrected version (CTTR) that is independent of sample size (Hess et al., 1984).

Hashtag density, overall (continuous). Hashtag density is a measure of the percentage of tweets that contain at least one hashtag. $Hd = \left(\frac{N_{\#}}{N}\right) \times 100$, where $N_{\#}$ = number of tweets with at least one hashtag, and N = total number of tweets (Shapp, 2014) for the overall tweet sample.

Hashtag type (categorical). Hashtags are either of type *tag* or of type *commentary* (hand-coded in the hashtag subset).

Hashtag language (categorical). Hashtags occur either in German or in English in German tweets (hand-coded in the hashtag subset).

Emoji density (continuous). Emoji density is a measure of the percentage of tweets that contain at least one emoji. $Ed = \left(\frac{N_e}{N}\right) \times 100$, where N_e = number of tweets with at least one emoji, and N = total number of tweets.

Sentiment score (continuous). Emojis are classified as either negative, neutral, or positive according to the continuous (-1, 0, +1) scale established by Novak et al. (2015). Here, the sentiment score per tweet is based on the emojis present in the tweet, using Novak et al.'s (2015b) sentiment scores to match the emoji. If there are two or more emojis present in a single tweet, the sentiment score for this tweet is the mean of those sentiment scores.

Positive emotion words (continuous). This is a measure of the percentage of positive emotion words.

Positive feeling words (continuous). This is a measure of the percentage of positive feeling words.

Negative emotion words (continuous). This is a measure of the percentage of negative emotion words.

Anger words (continuous). This is a measure of the percentage anger words.

Swear words (continuous). This is a measure of the percentage of swear words.

Modal verbs and words that express tentativeness (continuous). This is a measure of the percentage of modal verbs and words that express tentativeness.

Social concerns (continuous). This is a measure of the percentage of words pertaining to social concerns.

Family and friends (continuous). This is a measure of the percentage of words pertaining to family and friends.

Occupation (continuous). This is a measure of the percentage of words relating to occupation.

Job (continuous). This is a measure of the percentage of words relating to job.

Achievements (continuous). This is a measure of the percentage of words relating to achievements.

Sports (continuous). This is a measure of the percentage of words relating to sports.

Money (continuous). This is a measure of the percentage of words relating to money.

Formal *denn* 'because' conjunction (continuous). This is a measure to determine if the German tweets under investigation are more oral-like or lean towards a more written genre.

Informal *weil* 'because' conjunction (continuous). This is a measure to determine if the German tweets under investigation are more oral-like or lean towards a more written genre.

3.3 Language Processing Under a Psychological Lens

3.3.1 Linguistic Inquiry and Word Count

The sheer volume of data makes computerized, automated data analysis indispensable for researchers to be able to handle an ever increasing data load, and why Mehl (2006) refers to Linguistic Inquiry and Word Count (LIWC) as currently being the most-used, and best-validated software in psychological research for automated text analysis. In addition, Einspänner, Dang-Anh, and Thimm (2014) mention LIWC as one of the great tools for Twitter content analysis. The software, Linguistic Inquiry and Word Count (LIWC), was developed by psychologists, Pennebaker and King (Pennebaker, Booth, et al., 2007; Pennebaker & King, 1999) to facilitate measurements of language use as individual differences (using the Big Five trait taxonomy). Since, LIWC has been used in a multitude of studies to assess personality through language use (Pennebaker & King, 1999). "The social sciences have entered the age of data science, leveraging the unprecedented sources of written language that social media afford" (Schwartz et al., 2013, p. 1). The preceding quote aptly illustrates why Schwartz et al.(2013) mention LIWC as one of the most widely used language analysis tools. As a dictionary based analysis tool, LIWC utilizes a dictionary (~6,400 English words in the 2015 version) with 76 different word categories as a baseline, against which text input is compared and percentages for individual word and content categories are computed.

LIWC (Pennebaker, Booth, et al., 2007; Pennebaker & King, 1999) processes text input and computes the percentage of total words that match each of the dictionary categories. It operates with two central features: the processing unit, and the dictionaries. The processing unit (the software itself) opens and goes through text files word by word and then compares each word with the built-in dictionaries. Dictionaries, which are at the core of the program, refer to the collection of words that make up a dictionary. For example, the category of articles encompasses three words: "a," "an," and "the." Other categories, such as positive or negative emotion words, and words pertaining to power, and social relationships are more subjective. Initially, word candidates for each category were compiled from dictionaries, thesauruses, questionnaires, and lists by research assistants. Then, groups of three judges independently rated the words and determined if they were appropriate candidates for the overall word category in question. The rating process was repeated twice to ensure higher accuracy resulting in an agreement rating in between 93% and 100% (Pennebaker, Booth, et al., 2007; Pennebaker & King, 1999). After the initial dictionaries had been compiled in 1992 and 1994, significant revisions were implemented in 1997 and 2007 to improve accuracy further. Throughout this streamlining process, more than 100 million words were analyzed by the creators of the software (Pennebaker, Booth, et al., 2007; Tausczik & Pennebaker, 2010). Assume a text with 2,000 words, of which 150 are pronouns and are 84 positive emotion words is being analyzed. LIWC computes the percentages for pronouns (7.5%), and for positive emotion words (4.2%) respectively. Now, while percentages are a lot better for statistical analysis than frequencies, it goes without saying that more words are better than fewer words. LIWC's developers suggest a minimum of 50 words for successful analysis (LIWC, 2016). As will be seen below, once converted to raw text, the tweet sample becomes a word sample with the number of tweets being less crucial for successful analysis than the overall number of words and analysis happening at the word level.

It is important to mention a caveat at this point: nearly all text analysis programs rely on word counts and can therefore not account for context, irony, sarcasm, or even multiple word meanings. While this is not ideal, "small classification errors like multiple word meanings rarely impact the conclusions that can be drawn from the results because they are offset by the way that words are most commonly used by people" (LIWC, 2016).

With correlations ranging from .20 to .81, Pennebaker, Booth, and Francis (2007) proved the external validity of word categories to be included in LIWC (also see Back et al., 2010; Golbeck, Robles, Edmondson, et al., 2011; Golbeck, Robles, & Turner, 2011; Mairesse et al., 2007; Pennebaker & King, 1999; Qui et al., 2012; Schwartz et al., 2013; Tausczik & Pennebaker, 2010; Vazire & Mehl, 2008; M. Wolf et al., 2008; Yarkoni, 2010).

3.3.2 The German LIWC Dictionary.

Since LIWC is dictionary based, usage in a language other than English requires a dictionary in said language, which has to be tested for equivalence to the English dictionary, validity, and for robustness in terms of errors. In a twofold approach, Wolf et al. (2008) tested the equivalence of a German dictionary for LIWC and the robustness of the dictionary in terms of spelling errors. They found that both versions are equivalent with certain limitations. The content word categories are equivalent except for the "cause" and "space" categories. All other categories lie within a 90% confidence interval and can therefore be considered equivalent according to Wolf et al. The 2015-version of LIWC with the English LIWC-dictionary (Pennebaker, Boyd, Jordan, & Blackburn, 2015) is quite a bit more extensive than the original 1999/2000 LIWC dictionary, on which the German LIWC dictionary (German2001.dic) is based. Since no updated version of the German LIWC dictionary was available, the 2001-version was used for all analyses. The German 2001-LIWC dictionary currently comprises 7,598 words and is thus more extensive than the 2015 English LIWC dictionary with ~6,400 words.

3.4 Sample

The dataset in this study is twofold: data were gleaned from Twitter in the form of tweets (the Twitter API makes 3,200 of a user's most recent tweets available (Twitter Inc., 2017c)) and from participants' questionnaires to provide insight into their usage habits on Twitter and obtain more detailed demographic information. Data collection via the online questionnaire started on January 10, 2017 and ceased on May 25, 2017. The participants' tweets were collected via Twitter's API on May 25, 2017.

Since LIWC relies on a word-count approach rather than on a tweet count approach, users who had fewer than 50 unique words in their sample were excluded, see more on that below. As mentioned above, LIWC's developers suggest a minimum of 50 words for successful analysis (LIWC, 2016). The minimal number of tweets to get significant results is vague and previous research has used anything in between around 30,000 tweets (Golbeck, Robles, Edmondson, et al., 2011; Golbeck, Robles, & Turner, 2011) to several hundred million tweets (Qui et al., 2012). Since this study has two data sources, a smaller number was considered sufficient, especially because Germans tweet a lot less (Doyle, 2014; Eisenstein et al., 2014; Scheffler, 2014). For example, Qui et al. (2012) screened participants such that only those who posted more than 20 and fewer than 1000 tweets during a single month were included. While Qui et al. (2012) collected their tweets from multiple users over the period of one month through the Twitter streaming API, in this study, tweets were collected from the users directly on a single day. That means that for any given user, their most frequent 3,200 tweets were collected, since Germans do not tweet nearly as much as Americans for example, the collected tweets are likely to be an actual representative sample of German Twitter users (Scheffler, 2014). Thus, for some participants, it was possible to include literally all their tweets, starting from when they first joined Twitter, while for other, more prolific users, this was not possible, and the collected tweets only reached back until 2012, depending on their level of activity, for example. This approach has a definitive advantage over using the Twitter API to get a random sample of tweets and then using a German stop word list to filter out German tweets with keyword tracking (e.g. Scheffler, 2014), because the sample is already relatively clean in terms of user language.

The data thus consists of micro-corpora collected from individual users. This way, I was able to ascertain that tweets came from Germans exclusively regardless of where or in what language the tweet was made: in Germany or abroad, or in German or English. Furthermore, the data are a lot cleaner as the tweets will not come from a random sample created by Twitter, which could be skewed and contaminated by spam bots or companies advertising their products on Twitter instead of creating content for the Twitter-sphere. Instead they only consist of tweets coming from the actual participants.

As illustrated above, many studies have investigated the relationship of language and personality, mostly in a tightly contained university context. However, according to Krantz, "[n]o one has ever gotten a random sample in the lab" (as quoted in, Azar, 2000, p. 42), which is why the validity of web-based studies must also be considered against the backdrop of this methodology. Gosling et al. (2004) looked at six common preconceptions about web-based data: (1) Samples are not diverse, (2) the Internet is used by socially inept people, (3) collected data are affected by the presentation format, (4) web-based questionnaires are affected by non-serious responses, (5) findings are adversely affected by the anonymity on the web, and (6) web-questionnaire findings are inconsistent with findings from traditional methods. Their overall finding was that web-based samples are at least as diverse as samples from traditional methods, and that, although internet samples are not random samples of the population, obtaining 'real' random samples is not just an ‘online’ problem in the first place (Gosling, Vazire, Srivastava, & John, 2004). I suggest that, over ten years later, their findings have become even more important as the internet and social media have moved from desktops into the wholesale ubiquity of mobile devices, which legitimizes and corroborates the proposed sampling methods even more.

3.5 Participants

While Qui et al. (2012) used three different methods to recruit participants (snowball sampling, on-campus recruiting, and Amazon’s Mechanical Turk), in this study, snowball sampling was used as the main recruitment method. The participants are $N = 62$ German Twitter users (ages 18-45;[18] here, I follow Küfner and Back’s (2010) age bracket), who filled out a questionnaire on demographic information and their Twitter usage. According to the Central Limit Theorem (CLT), as samples get larger (usually $N > 30$), the sampling distribution approximates a normal distribution with a mean that is equal to the population mean (Field, Miles, & Field, 2013). While the actual definition of a ‘large’ sample is contentious, Hogg, Tanis, and Zimmerman (2015) suggest a sample size of “*N* greater than 25 or 30,” (p. 202). However, Cohen (1990) claims that 30 is not enough. I thus proceeded on the assumption that the “[s]ample size is never large enough. […] [and] inferential needs increase with [your] sample size” (Gelman & Hill, 2007, p. 438). Sixty-two participants, however, seems to be well within that somewhat vague definition of what is considered a large enough sample. Equal group size for both the male and the female groups was almost achieved (male = 29; female = 33). On this note, it has been thoroughly established that unequal group sizes do not pose a big problem for most statistical models following a maximum likelihood approach (e.g. Linear Models,

[18] Age is nearly normally distributed.

Generalized Linear Models, Multilevel Hierarchical Mixed Models, GAMs, ANOVAs) as the statistical power usually is only decreased by negligible values. Especially if the difference in n is fairly small, as is the case here, and the variances in both groups are fairly homoscedastic (Maxwell & Delaney, 2004).

3.6 Recruitment

To recruit participants, the author disseminated the link to the Qualtrics survey/questionnaire on his own Twitter, Facebook, and LinkedIn accounts/timeline (as well as sent direct messages to his Facebook and Twitter network), and posted it on the website, Surveycircle.com,[19] to increase the number of participants through snowball sampling. This method was used successfully in a similar context by Pennebaker et al. (2007) and Qui et al. (2012). Participants were also asked to post the survey link on their own timeline and/or send the survey through direct messages to their respective social media networks/connections on Twitter and/or Facebook/LinkedIn to reach more participants and cast a wider demographic net. The participants were, however, asked not to post the survey on somebody else's (e.g. one of their friend's) timeline (Twitter/Facebook/LinkedIn) directly to ensure that the questionnaire did not spam someone's timeline and potentially violate privacy restrictions imposed by Twitter/Facebook/LinkedIn, and to make sure the recipients of the study did, in fact, belong to the extended social network of the researcher and its first connections. This approach ensured that not only a limited student population, represening only a limited age range and educational status, but also people outside of academia, of different ages, and with different educational

[19] Surveycircle.com lets PIs post their surveys online, which other PIs/survey managers can fill out. On a quid-pro-quo basis, this facilitates the recruitment process by helping fellow researchers and getting help [participants] in return (Surveycircle, 2017, March 15).

backgrounds were sampled. The participants also provided their Twitter handle/ID, so their tweets could be collected. Users who had fewer than 50 unique words in their sample were excluded. No upper bound was used as an exclusion criterion, as this leads to more robust models, and analyses become more accurate with more data. This was done to ensure that there was enough text data from participants to be analyzed in LIWC. To reiterate, LIWC's developers suggest a minimum of 50 words for successful analysis (LIWC, 2016).

Overall length of a user's membership to Twitter was not considered as an exclusion criterion, as the usage rate of any social medium is a highly individual factor, which can provide insights into the users' Twitter habits.

3.7 Data Collection

3.7.1 Ethical and Legal Considerations

The information and content posted on Twitter is routinely considered public domain, unlike Facebook, for instance, where stricter ethical rules apply, given its partially non-public nature (D'Arcy & Young, 2012). While Twitter research is still subject to legal uncertainty to some extent (Beurskens, 2014), Twitter's relationship to its users is regulated in its "Terms of Service" (Twitter Inc., 2017d). Regulations for developers (including the API) can be found in Twitter's "Developer Agreement and Policy" (Twitter Inc., 2017b) and Twitter's "Privacy Policy" (Twitter Inc., 2017e), which controls and regulates intended use of data collected from/on Twitter. Although open data bases would be advantageous for transparent and reproducible research, Twitter has repeatedly tried to ban, and disapproves of, so-called "shadow data bases." Thus, a researcher may not grant other researchers access to their data. Once data leave Twitter's server however, it cannot stop researchers from archiving, sharing, or reusing data they obtain from Twitter. However, Twitter can redesign its API at any point preventing people from consistently collecting data (Beurskens, 2014). Keeping in mind that "while users voluntarily publish their everyday activities and opinions [on Twitter], they do not automatically also agree to the use of such data in any way imaginable" (Beurskens, 2014, p. 131), data collection on Twitter was deemed ethically acceptable as was the data collection on Qualtrics with an online questionnaire since the data were stored securely and remain with the researcher to be published mostly only in agglomerated form making it nearly impossible to reverse-engineer the identity of the participants (exempt are tweet samples in the study). Since the participants have public profiles (one of the inclusion criteria), there were no issues in terms of privacy infringements as their tweets are publicly available even to people who do not have a Twitter account at all. This is in alignment with current research ethics as recommended by the Association of Internet Researchers (AoIR) (esp. Markham & Buchanan, 2012, pp. 4-10). The proposal for this study was submitted to the Institutional Review Board (IRB) at Ball State University to meet institutional practices and research ethics requirements. The study was approved by the Ball State Institutional Review Board (Protocol(s): 979954-1/2/3).

3.7.2 Questionnaire

Participants were asked to fill out a questionnaire (see Appendices A and B, pp. 249-257) which gleaned information about the above-mentioned sociological variables, such as age, gender, educational background, level of education, geographic location, their Twitter behavior (adapted from, Hughes, Rowe, Batey, & Lee, 2012) and their personality based on the short, 10-item version of the Big Five personality inventory, the BFI-10 (Gosling et al., 2003; Rammstedt

& John, 2007; Rammstedt et al., 2012), which has been used successfully in similar contexts (Küfner et al., 2010).

3.7.3 Twitter Data Collection

The Twitter API[20] was used to collect the 3,200 (up to) most recent tweets from the participants to build individual corpora for each participant and a combined corpus, which could then be processed with natural language processing tools (LIWC). Research has shown that writing patterns are relatively stable and consistent over time, and over the life-span for that matter (Pennebaker & King, 1999; Pennebaker & Stone, 2003), which means that it does not matter if a tweet is one hour, 20 days, or two years old. The tools needed for this part of the data collection are relatively simple: a personal computer (Apple™ MacBook Pro 15", 2.4 GHz Intel Core 2 Duo, 8GB RAM) and the R-package, **`twitteR`**, version 1.1.9 (Gentry, 2016), which does the legwork of the tweet collection and puts participants' tweets in a data frame (i.e. spread sheet) in R used for further processing. This method streamlines the data collection process as no intermediary files are being produced — everything is handled in one programming environment.

With this twofold approach to data collection, Twitter data were supplemented by more in-depth demographic information, such as gender, the exact age, zip codes, relationship and education, as well as Twitter measures, such as minutes spent on Twitter per day. In addition, the observer's paradox (Labov, 1972) did not skew the data, as the tweets constitute natural language data and thus suitable for (socio)linguistic and personality analysis. In fact, Twitter comes close

[20] Application program interface or application programming interface. An API is a defined set of routines and protocols that allows communication in between two (or more) software components. APIs are used by developers to integrate information from different sources in dashboards for example (Application-programming-interface, 2017).

to near face-to-face interaction (Naaman et al., 2010). Twitter accounts have previously been linked to other data sources to gain a deeper understanding of users' tweet and online behavior, for example Burger et al. (2011), who linked 184,000 Twitter accounts to blog profiles with demographic information.

3.8 Procedures

3.8.1 Questionnaires

The data from the Qualtrics questionnaires were stored in csv-file format on Dropbox and in an agglomerated format as a data frame on GitHub. Before any data were processed, the questionnaires were hand-checked to see if (1) the participant entered a valid Twitter account/handle (verified on Twitter), and (2) if the participants filled out all parts of the questionnaire. Only participants with active German Twitter accounts and completed questionnaires were included. The csv-file was then processed with R/RStudio (R Development Core Team, 2017; RStudio Team, 2017) to clean and re-format data for statistical analysis, and to extract the variables to be included in the statistical models. An agglomerated dataset was built containing the demographic information for all participants in a single data frame (spread sheet). Additionally, the Qualtrics survey provided participants' scores on the BFI-10 personality inventory. The mean scores for each personality trait were calculated in R for each participant individually. This way, each participant received a score for each of the five personality traits, which were added to the agglomerated data set. That way, I was able to use personality measurement as covariates in the statistical analysis.

3.8.2 Twitter Data Processing

The first step in Twitter/tweet processing was to check individual participant accounts for potential Twitterbots. A bot is an automated program that sends tweets automatically or follows users automatically. They can also automatically retweet and reply in an @reply (Chu, Gianvecchio, Wang, & Jajodia, 2012; Hill, 2012, August 9). Twitterbots, which automatically disseminate tweets, are not a problem in themselves (Mowbray, 2014). The problem for data analysis is tied to the numbers: as early as 2009, 24% of tweets were said to have come from bots (Cashmore, 2009, August 6), with Twitter admitting in 2014 that around 8.5% (or 23 million active users) of all user accounts were bots (Seward, 2014, August 11). On a relatively small scale, bot detection is relatively easy, as individual user accounts can be hand-checked (Chu et al., 2012), which is what I did in this study. Conversely, tweet automation, as it is offered through online applications such as Twittimer (2017), which allow the user to schedule tweets, are not a problem because the users' language is going to be the same — the crux of the matter in the analysis at hand.

Once participants' Twitter accounts were verified and incomplete or inactive Twitter accounts excluded, the above-mentioned R-scripts (available upon request from the author) were used to (1) collect all the tweets from individual participants. Then, user-specific tweets were counted to get an overall estimate, leading to the exclusion criterion (also see above) of users, who had fewer than 50 unique words in their tweets (the minimum number of words for LIWC to produce meaningful results). No upper bound was used as an exclusion criterion to follow the paradigm, the more data, the better. (2) To test the language hypotheses, participants' tweets were first 'cleaned' (see below).

Any retweets (RTs) were excluded in the data collection as RTs do not constitute original and user-specific tweets and, thus language use. That way, the resulting tweet-set exclusively consists of tweets written by the participant. @-mentions, however, remained in the tweet dataset because they were produced by the participant. (3) Since Germans do also tweet in English, the R package, **`cldr`**, version 1.1.0 (McCandless & Sanford, 2013a),[21] was used to extract the German tweets, which is what I focused on in this study. (4) From here, three operations were carried out: (i) German tweets were processed with another R-script to extract only the normalized (see below) tweet text (tweet-body) to be saved in individual text-files (.txt), since this is the data-portion and file-format needed for analysis with LIWC. LIWC then provided overall statistics, such as total number of words, and calculated the percentages for the dependent variables, such as percentage of positive emotion words, and negative emotion words. (ii) Concurrently, the tweets in the German tweet data set were labeled (contains at least one hashtag: yes/no) for the computation of the overall hashtag density. Out of all the tweets containing a hashtag, a randomized sample was drawn for further analysis of hashtags (tag/comment), and (iii) an R-script filtered out the emojis in the German tweet dataset to calculate emoji measures, i.e. emoji density and sentiment scores.

3.8.3 Twitter Text Normalization

Since most social media texts are very noisy, i.e. they contain hashtags, @-signs, repetitive use of letters, and other symbols, noisy text normalization (NTN) was applied to

[21] The R package, **`cldr`**, is a language detection tool that uses Google Chrome's language detector/identifier. In its essence, it is a wrapper function for the C++ Compact Language Detection library, which is based on the Chromium project (McCandless & Sanford, 2013a, 2013b). It is no longer available on R's CRAN-server, but can still be installed using R's **`devtools`**-package: **`devtools::install_version("cldr",version="1.1.0")`**.

reduce the noise and normalize, i.e. clean, the tweets (A. Clark, 2003; E. Clark & Araki, 2011; Sidarenka, Scheffler, & Stede, 2013). As any kind of text is a rather unstructured collection of words for a computer (Feinerer, Hornik, & Meyer, 2008), NTN is necessary not least to make a computer 'understand' the text input. NTN is not always mentioned in studies pertaining to social media texts, because it is either considered a given, since the hypotheses being tested did not require any textual pre-processing (e.g. Scheffler, 2014), or, in a worst case scenario because it was simply not done potentially at the cost of accuracy (e.g. Qui et al., 2012). Here, I mention it for the sake of completeness, and, more importantly, because structured pre-processing (with regular expressions[22]) was implemented and applied to the tweets at hand to make them readable and ready to be processed by LIWC. In addition, this yielded more accurate results during the tokenizing/tagging process, and ultimately, in the calculation of lexical diversity measures (Carroll's CTTR and Yule's K). With Twitter in particular, there are many Twitter-specific phenomena (TSP), such as @-mentions, and #-hashtags (Kaufmann & Kalita, 2010), which makes normalizing tweets a non-trivial task due to the high variability of symbols in the text. Additionally, there are other TSPs, such as links (email addresses and hyperlinks) and emojis that need to be dealt with to clean the raw text. I chose to delete links and emojis in the text-only corpus while, in the emoji-corpus, everything was retained as the emoji analysis does not require the tweet (text) to be particularly clean. To deal with meta-characters, Kaufmann and Kalita (2010) suggest syntactic disambiguation as both the @-mention and the hashtag can be part of the syntax.

This entailed the deletion of the meta-characters @ and # for the text only corpus, as discussed below. However, hashtags were only deleted at the end of a tweet, and not if they were

[22] A regular expression (RE), or regex, is used to specify a text pattern with a sequence of characters in computer science. REs are often used to find or replace/manipulate strings of text (Jurafsky & Martin, 2009).

part of the syntax as a part of speech. In the latter case, the #-metacharacter was deleted at the beginning of the word. In addition, punctuation, as well as leading and trailing white spaces, were also deleted. These deletions do not affect the text itself, as any actual lexical items to be analyzed were retained. Clark and Araki (2011) also mention emoticons (here: emojis), which can distort text analysis. They were thus also excluded in the text-only corpus. Further, Sidarenka et al. (2013) point out that words in social media texts often have reduplicated vowels, and to a lesser extent consonants, to express nuances of intonation not otherwise available in written communication, e.g. *Hilfeee* 'help,' or *süüüßßß* 'cute.' When implementing an algorithm to clean tweets with regular expressions, one has to keep in mind that, in German, the double consonants, or *Doppelkonsonanten* (b, d, f, g, l, m, n, p, r, s, and t) must not be reduced to only one consonant in inter-vocalic position (Fuhrhop & Barghorn, 2012), e.g. *Hallllllooooo* 'hello' ≠ *Haloooo*/*Halo*. The eszett (ß) does not occur in double form and can therefore be reduced to only one instance. Double vowels are relatively rare in German orthography (a > aa, e.g. *Haar* 'hair,' e > ee, e.g. *Meer* 'sea,' and o > oo, e.g. *Boot* 'boat') and were treated as a limited list of words to be analyzed, i.e. they were included as exceptions in the text-normalization algorithm. Overall, there are only six German words spelled with <oo>, *Boot* 'boat,' *doof* 'dumb,' *Koog* 'polder,' *Moor* 'swamp,' *Moos* 'moss,' and *Zoo* 'zoo.' Words with <aa>, *Aal* 'eel,' *Haar* 'hair,' *Saal* 'hall,' *Saat* 'seed,' *Staat* 'state,' and *Waage* 'scale,' and <ee>, *Beere* 'berry,' *Beet* 'patch,' *Fee* 'fairy,' *Heer* 'army,' *Klee* 'clover,' *leer* 'empty,' *Meer* 'sea,' *Schnee* 'snow,' *See* 'lake,' *Seele* 'soul,' *Speer* 'spear,' and *Teer* 'tar' are barely more frequent (Westermann, 2017). In any of these rare cases, the vowel in the word must not be reduced to only one vowel. Also, an elongation at the end of a word as in *sooooo* 'so' below (Example 2a) must not result in *soo*, which had to be factored in when writing the regular-expressions for tweet-normalization. In addition, German, just like English, is a language with many regional variants; variants that do not only manifest themselves between individual states, but also within states with individual cities sometimes having their own dialectal features (e.g. Berlin, which is a special case as a city-state) (Berend, 2005). As German pronunciation, and thus phonological variation, is closely tied to its orthography, and people usually do not exclusively use Standard High German like many people living in and around Hannover (Cordes, 1983; König, 2005), I factored in as many of the more common regional variants as possible (esp. Berend, 2005, pp. 149-152) by writing specific regular expressions to filter out intricacies such as word final spirantization of plosives following a vowel as in *Tag* [tag] 'day' > *Tach* [tax], *sag* [sag] 'say' > *sach* [sax], or word final deletions following a voiceless fricative as in *nicht* [nɪçt] 'not' > *nich* [nɪç]. Another great example is the Berlin-variant of the first person personal pronoun *ich* [ɪç] 'I,' which is commonly written and pronounced as *ick* [ɪk] or *icke* [ɪkə]. Research has well established and attempted to address the issue of a 'standard' language, which is particularly difficult in English, for instance. However, German is no exception and the question what linguistic features exactly constitute the German language is and who speaks it has not been answered conclusively, not least because of the exceptional difficulty of describing the continuum of varieties that exist along the standard-dialect-axis (Barbour & Stevenson, 1998; Lameli, 2004). In Germany, there exists a large number of fluid dialectal regions with smooth linguistic boundaries, which transition into each other, so that any one specific standard German language is hard to capture, maybe even intangible - Standard High German (SHG) is spoken in and around Hannover (with exceptions) (König, 2005). Like in English, there are many regional variants, with some cities even having their own dialectal features (e.g. Berlin), which further confirms the theory that discrete linguistic boundaries are a relatively rare phenomenon (Barbour & Stevenson, 1998). I mention this for good measure and because, even though I attempted to implement a very thorough text processing/cleaning algorithm capturing and covering the broadest and most common dialectal features or differences (e.g *Tag* [tag] ‘day’ > *Tach* [tax] (Berlin) vs. *Tag* [tag] ‘day’ > *Dog* [dåg] (Bavaria), it is nearly impossible to capture all German linguistic features with all their dialectal intricacies in pronunciation (often mirrored in the orthography) and individual lexical variants for certain words when looking at the broader picture (German) and not focusing on a particular dialect area (e.g. Bavaria, or one of its sub-regions). The fact that some researchers argue that certain dialects might even be a language in their own right (e.g. Bavarian, H. U. Schmid, 2012) muddles the waters further, if anything. This invariably introduces as certain amount of error that needs to be kept in mind when running the text cleaning algorithm. As additional measures of formality, punctuation marks (‘.’, ‘,’, ‘;’, ‘:’, ‘?’, ‘!’) were retained in the cleaned tweet samples to be analyzed by LIWC (Thayer, Evans, McBride, Queen, & Spyridakis, 2010). Ultimately, the made-up tweet in Example (1a) becomes (1b) after noisy text normalization:

(1a) @abc Das #emoji ist sooooo süüüßßß! ☀️ https://www.emoji.com #cute #mussichhaben

The #emoji is sooooo cuuuuute! ☀️ https://www.emoji.com #cute #ihavetohaveit

(1b) Das emoji ist so süß!

The text in Example (1b) now has a form, which not only reduces heterogeneity among all tweets, but that can readily be analyzed by LIWC so that the words can be assigned to their word categories more accurately and lexical diversity measures (Carroll’s CTTR and Yule’K) can be computed more accurately with the R **koRpus**-package (Michalke, 2017a, 2017b). Figure 3 illustrates the entire data processing pipeline from start to finish.

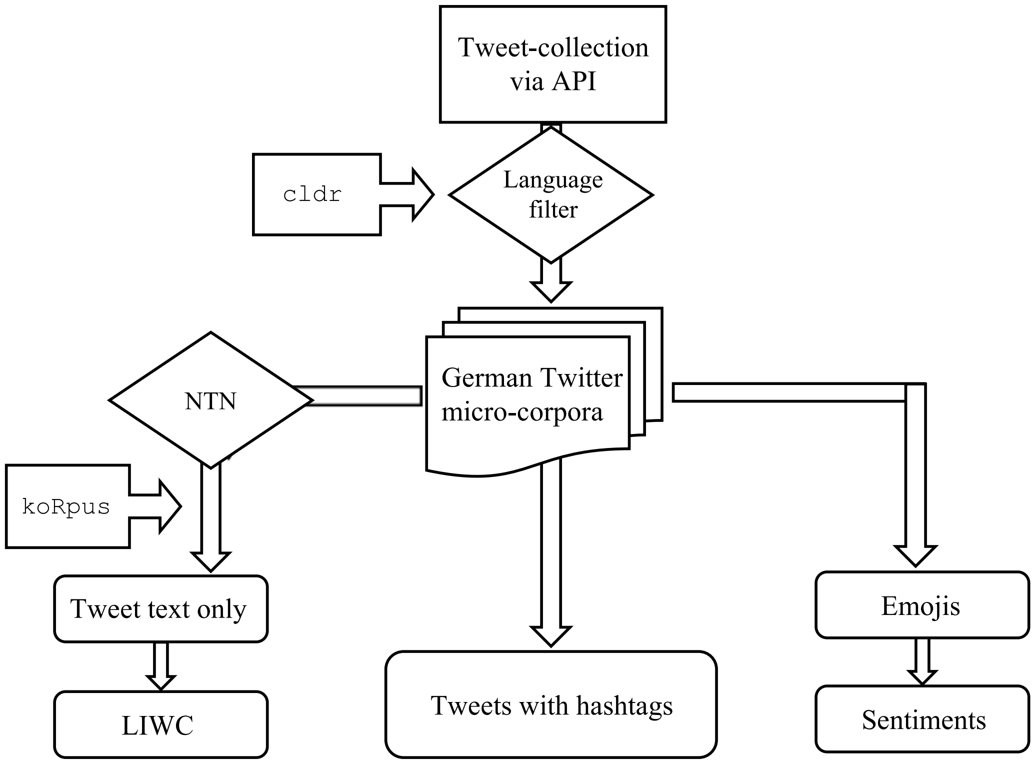

Figure 3. Tweet Processing Pipeline.

3.8.4 Hashtag Hand-Coding

Testing the hypotheses pertaining to hashtag use and gender is a little trickier because individual hashtags are not coded in the tweet sample and automated coding is impossible for my purposes. This is why samples of German hashtags were looked at to answer the question of whether females do in fact use more commentary hashtags compared to males, who purportedly use more tag hashtags (Shapp, 2014). Furthermore, I investigated how and if the "hashtag"-language (German/English) influenced the hashtag density (tag hashtags vs. commentary hashtags and hashtags overall). To do this, I extracted a subset of only tweets containing hashtags (*n* = 8,105), from which I drew a random subset (*n* = 1,621), which follows Shapp's (2014) approach. Shapp's (2014) sample of tweets with hashtags was 1,072, (with 1,633 hashtags overall), which yielded statistically significant results in her General Linear Mixed Effects Regression (**`glmer`**) model, which, adding random effects for participants as well as random slopes for position, explains away some of the random error. My hand-coded subset contains 1,621 tweets with 2,666 hashtags, which is bigger than Shapp's (2014) sample by a factor of ~1.63 — statistically significant results were thus expected. Since averaging the frequency of hashtags per participant does not yield any real insights because it is prone to be unevenly distributed, hashtag densities were computed for every participant, i.e. the number of tweets containing a hashtag was divided by the total number of tweets multiplied by 100. For tag and commentary hashtag densities, the number of tweets containing tag hashtags from the subset was divided by the total number of tweets containing a hashtag in the subset multiplied by 100 and, mutatis mutandis, the same was done for commentary hashtags. By reading every tweet in the subset, the hashtags were hand-coded for type of hashtag (t = tag; c = commentary), language (de = German; en = English), and position (b = beginning; m = middle; e = end) in the tweet. Correspondingly, Tsur and Rappoport (2012) call these positions *prefix*, *infix*, and *suffix* respectively, given their syntactic placement in the tweet. Coding hashtags that way follows Shapp's (2014) approach and seems to be the most failsafe way to accurately tease apart the hashtag categories, because generally, hand-coding is more accurate given the importance of the context for the type of hashtag, i.e. one that cannot easily be factored into a computerized algorithm automatically detecting hashtags. Table 1 below shows how the tweet-subset was structured for hashtag hand-coding: one row does not correspond to one tweet. Here, one row corresponds to one hashtag (this is indicated by the **`tweet_id`**). Based on the tidy data paradigm (Wickham, 2014), this method follows the tidy text paradigm (Silge & Robinson, 2017), in which there is one row per token in a given table; a token is commonly defined as a meaningful unit of text to be analyzed (Silge & Robinson, 2017).

Table 1: *Hashtag Template Data-Frame for the Tweet-Subset*

part_id	tweet_id	#_type	lang	pos
1	1	t	de	m
1	1	t	en	e
1	2	c	de	e
…	…	…	…	…
2	1	c	de	e
2	2	t	en	b
2	3	t	en	e
…	…	…	…	…
3	1	t	de	m

Hashtags were assumed to be of the *tag*-kind by default. If hashtags have ambient affiliation (Zappavigna, 2011), it is reasonable to assume that someone else would search for such a hashtag, making it a *tag*-hashtag used to index/organize a tweet, and thus connecting it to a discussion or online 'community.'

Many hashtags that tag people, events, companies, TV-shows, brands, etc. fall within the *tag*-category. When hashtags describe context, they are still considered tags since they nevertheless describe a specific tweet topic. Unlike *tag*-hashtags, *commentary*-hashtags do not highlight a topic. Rather, they are self-contained messages and as such subjective by nature. Thus, again following Shapp's (2014) framework, if a hashtag adds additional meaning to the content of the tweet, or if it is part of the semantic content of the tweet itself, it will be coded as commentary, see Example (2) for both hashtag types (**`part_id = 4`**):

(2) Stolz ist ein schlechter Berater. #word[c] #spruchdestages[t] #instaquotes[t] #erkenntnis[t] #impuls[t]

Pride is a bad advisor. #word #quoteoftheday #instaquote #insight #impulse

In Example (2), the hashtag, #word was coded as *commentary* [c], because it adds additional, subjective content to the tweet: the English #word refers to 'word,' used heavily in hip-hop culture in the late 80s and 90s, to lend weight to a statement.[23] Here, it adds weight to the tweet content and its truth value. The hashtag, #spruchdestages, has its own phrasal structure making it more complex and thus more likely to be a *commentary* tag. Here, however, I argue that #spruchdestages (#quoteoftheday) is a *tag*-hashtag connecting the tweet to other quotes of the day in the Twittersphere and defining the topic of the tweet. Topic-defining hashtags usually occur in suffix or end position (Zappavigna, 2015), which is the case in the example. This also illustrates how vague the distinction can be. Syntactic/phrasal complexity cannot be used exclusively to determine the hashtag type. The hashtags, #instaquote (en) and #impuls (de), are more obvious *tags* in that they connect the tweet to a larger audience on social media (Instagram in this case). The hashtag, #erkenntnis 'insight' is more ambiguous, either adding a comment to the tweet body, or connecting it to the Twittersphere to make it searchable for other users looking for insights. Again, tags are assumed to be the default, so I coded #erkenntnis as a *tag* to maintain a systematic coding process. Example (2) quite nicely shows how Germans tend to use

[23] Etymologically, 'my word is my bond' goes back to the Latin phrase *dictum meum pactum*, which in 1801 was adopted as their motto by the London Stock Exchange, where deals are made without written documents. It means *I'm speaking the truth*/*I'll keep my promise*. Part of the rich hip-hop history, it has since been shortened to 'word is bond' and further to 'word,' *truth*/*promise* (Mywordismybond[Def.1], 2017; Waldman, 2016, July 19; Word[Def.3], 2017).

hashtags quite a bit: there are five hashtags, exceeding Twitter's suggested limit by three hashtags (Twitter Inc., 2017g), for an utterance (tweet) of 'only' five words.

To reiterate, the goal here is not a discourse analysis, investigating users' motivations for the use of either or both kinds of hashtags. Nor is it an investigation of the syntactic structure of hashtags or their *prefix*, *infix*, or *suffix* roles. Rather, the two-fold distinction is crucial for hand-coding hashtags in the participants' samples as either tag or commentary hashtags to quantify individual hashtag use and to draw conclusions about gender and/or personality effects.

3.8.5 Emojis

An R-script based in part on Peterka-Bonetta's (2017b) blog post and GitHub repository (2017a) was used in conjunction with an 'emoji_decoder'[24] (Suárez-Colmenares, 2017) based on (Whitlock, 2017) to extract the emojis in question and retrieve the overall sentiment scores (Novak, Smailović, Sluban, & Mozetič, 2015a; Novak et al., 2015b) for participants and both genders. This way, the overall sentiment score (-1, 0, +1) could be added to the emoji density for each participant and each of the participants' tweets. Initially, tweets were converted from bytes (\xF0\x9F\x98\x8A) to a format the R-scripts could work with, see Examples (3a) and (3b) (**`part_id = 4`**):

(3a) "Heute lasse ich los, was mich beengt, beschwert, mir im Weg steht (... ich selbst inklusive ☺️)…. https://t.co/9TIFOAD5zq"

[24] This 'emoji_decoder' comprises 1442 (minus the skin tone variations) of the most current emojis including their descriptions, utf-8 encodings, and unicodes.

Today, I’m letting go of things that inhibit, encumber, or stand in my way (...myself included ☺️) […]

(3b) "Heute lasse ich los, was mich beengt, beschwert, mir im Weg steht (... ich selbst inklusive <e2><98><ba><ef><b8><8f>). <e2><80><a6> https://t.co/9TIFOAD5zq"

Example (3b) shows the utf-8 encoding for the emoji, <e2><98><ba><ef><b8><8f>. This encoding allowed me to then match the emojis with the emoji list and the sentiment scores. To reiterate the caveat from Chapter 2: Novak et al. (2015b) found that tweets with emojis skew positive with a mean sentiment score of + 0.365, which must be considered interpreting results. This caveat notwithstanding, sentiment scores are still a better indicator of emoji sentiment than a fixed list. In addition, using sentiment scores instead of a predefined list of positive and negative emojis allowed for more variation as dictated by the data and, ultimately, more accurate results.

3.9 Statistical Considerations

Since LIWC outputs percentages of individual word categories that can be correlated to personality measurements (Big Five scores), and I am not trying to predict personality scores from LIWC’s output, there was no issue with circularity,[25] which would diminish most model’s

[25] Circularity, sometimes also referred to as “double-dipping,” is “the use of the same data set for selection and selective analysis” (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009, p. 535). The problem is that this can not only distort descriptive statistics, but can also result in invalid statistical inferences (Kriegeskorte et al., 2009). Circularity should not be confused with (multi)collinearity. Collinearity refers to a situation in which two or more predictors in a multiple regression model are correlated. This can

statistical power. Furthermore, as mentioned above, unequal group size (male/female) did not turn out to be a problem for any of the statistical models since both groups only differ by four participants, male = 29; female = 33. The models to be used in analysis follow a maximum likelihood approach (e.g. Linear Models, Generalized Linear Models, Multilevel Hierarchical Mixed Models, GAMs, ANOVAs), which means that their statistical power usually is only decreased minimally if the difference in group size is fairly small and the variances in both groups are homoscedastic (Maxwell & Delaney, 2004). For model building, several considerations are important: (1) to build a good model (large effect size) with statistically significant results, around ten participants (observations) per predictor are usually recommended (Tabachnick & Fidell, 2013). If there is a large number of predictors, variable reduction techniques such as the Elastic Net or the LASSO (Zou & Hastie, 2005), or principal component analysis (PCA) are usually used to prevent overfitting[26] and thus reducing statistical power (Babyak, 2004). Usually, most researchers follow a parsimonious model building approach, which is built around the notion of Occam's Razor, i.e. a model should be as simple as possible but not simpler (Caffo, 2015). However, Andrew Gelman (2004) argues against parsimony for its own sake in his blog: "[…] if you can approximate reality with just a few parameters, fine. If you can use more parameters to fold in more information, that's even better" (Gelman, 2004, December 10). He further argues that while simpler models are easier to fit and take less effort to understand, we should not lend ourselves to the illusion that they are better than more

ultimately lead to inflated variances and thus a large effect size (R^2) even though none of the predictors are significant (R. M. O'Brien, 2007).

[26] When overfitting occurs, the model in question is too complex for the dataset being investigated; i.e. it fits the random noise in the data too well and would thus not work with another dataset that has its own random noise. An overfit model does not work with a new dataset and can not be used to model the entire population (Frost, 2015, September 3). Put another way, if too many unknowns are estimated, too many predictors are included in the model, and there might be complex interactions between the predictor and the outcome variable that only exist in the sample, but not in the population (Babyak, 2004).

complicated ones (Gelman, 2004, December 10). Given the somewhat contentious debate on the issue, I attempted to bridge both polar ends of the spectrum by fitting a simple model when possible, not shying away from more complicated and complex models for the sake of easiness.

3.9.1 Generalized Additive Models

3.9.1.1 The Linear Framework: Linear Models and Generalized Linear Models

Important tools in inferential statistics, likelihood-based regression models presume a likelihood for a given response variable (y), modeling the mean as a linear function of a set of linear (or other parametric) covariates ($X_1, X_2, \ldots, X_p$). Using maximum likelihood estimation (MLE), we get the parameters of the linear function (T. J. Hastie & Tibshirani, 1986). In a linear regression model, we investigate/predict an outcome variable, which we believe to be a function of (an)other variable(s). The outcome variable (y) is assumed to be normally distributed with mean μ and variance σ^2, $y \sim N(\mu, \sigma^2)$. The Xs are the predictors/covariates, represented in Equation 1:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_p X_p \tag{1}$$

The coefficients (β) are obtained using ordinary least squares (OLS); summing these coefficients results in the linear predictor, which also directly yields the estimated fitted values (M. Clark, 2016, June 26; Field et al., 2013). The problem with linear models is that they can be limiting in terms of their scope, i.e. ability to model polynomial or wiggly curves. If there is a dichotomous (binary) outcome variable, a linear model does not work and a switch to a generalized linear model (GLM) is necessary. A GLM allows for other types of distributions, such as binomial, or Poisson, adding a link function, *g()*, to the equation (e.g. linear regression: link = identity; logistic regression: link = logit). Said link function relates the expected values to a linear predictor still assuming normal distribution for the outcome variable and homoscedasticity for all observations (M. Clark, 2016, June 26; Gelman & Hill, 2007; Larsen, 2015, July 30), see Equation 2:

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \tag{2}$$

I offer this comparison not to diminish linear models' efficacy or dispute their usefulness in many statistical analyses and research settings in any way, but just to show their limitations when it comes to handling non-linearity, and the investigation of patterns that often get overlooked because a general additive model would have been a better fit and a linear model would have fit a straight line through a more complex underlying pattern.

3.9.1.2 The Generalized Additive Model (GAM)

The Generalized Additive Model was invented by Trevor Hastie and Robert Tibshirani (1986). Although Generalized Additive Models (GAMs) have been in existence for a while, data scientists and researchers in many disciplines are not making a lot of use of this statistical technique, which some have even labeled "the predictive modeling silver bullet" (Larsen, 2015, July 30, p. 1), and called for more extended applications. While GAMs are certainly not a silver bullet, just like most statistical techniques, they lend themselves to analyzing natural data for one simple reason: natural relationships are often non-linear, which can be a problem for standard linear regression models (SLiM) estimated via ordinary least squares (OLS). This makes GAMs especially applicable to linguistic/language data, as relationships between predictors can often be non-linear. Winter and Wieling (2016) showed exactly this in their analysis of linguistic change over time for instance. Assuming basic familiarity with linear regression and generalized linear models (GLMs, e.g. logit), I will present a short introduction to GAMs and why they were employed in this study, depending on the hypothesis to be tested.

Larsen (2015, July 30), Clark (2016, June 26), and Wood (2006) praise GAMs for several reasons: (1) they are easy to interpret, (2) oft-overlooked hidden patterns in the data can be revealed, and (3) we can avoid/reduce overfitting by regularizing (penalizing) the predictor(s). Simply put, "[r]elationships between the individual predictors and the dependent variable follow smooth patterns that can be linear or non-linear" (Larsen, 2015, July 30, p. 1). Another advantage is that GAMs support various link functions like GLMs, such as the logit link for a dichotomous dependent variable (T. J. Hastie & Tibshirani, 1986; Larsen, 2015, July 30).

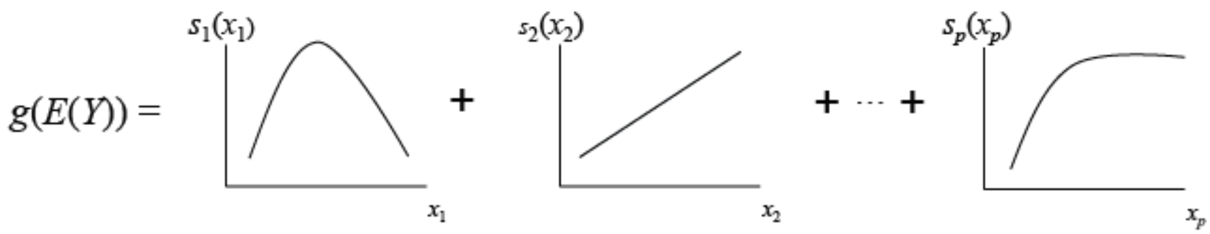

Figure 4. Graphical Representation of Possible GAM Structure (Larsen, 2015, July 30, p. 2).

Figure 4 shows how a GAM captures linear relationships (middle), just as well as polynomial relationships (left and right). The additive model extends the linear regression by adding a smoothing function, which denotes the impact of the predictors on the dependent variable showing underlying patterns, which can be non-linear (T. J. Hastie & Tibshirani, 1986, 1990; Larsen, 2015, July 30), see Equation 3 and compare to Figure 4 above:

$$g\big(E(y)\big) = \beta_0 + \sum_{i=1}^{p} f\,(X_i) \tag{3}$$

Here, y is the dependent variable, E(y) is the expected value, with g(y) being the link function that relates the outcome to the predictors X_1, …, X_p. Beta zero (β_0) is the intercept (like in a SLiM), and $f(X_1)$, …, $f(X_p)$ are smooth nonparametric functions, which are estimated iteratively and automatically during model estimation (T. J. Hastie & Tibshirani, 1986, 1990; Larsen, 2015, July 30; Wood, 2006).

These smooth functions $f(\cdot)$ can best be explained by invoking this simple example from Wood (2006, p. 120) in Equation 4:

$$y_i = f(X_i) + \varepsilon_i \tag{4}$$

Here, y_i is the dependent variable, X_i is a predictor variable, f is a smooth function, and ε_i, the error term, consists of normally distributed random variables, $N(0, \sigma^2)$. To estimate f, (4) has to become a linear model. By choosing a *basis*, or *basis function*, *b*, which is definitively known, we select a space of which f is an element (Wood, 2006):

$$f(X) = \sum_{i=1}^{p} b_i\,(X_i)\,\beta_i \tag{5}$$

Equation 5 illustrates the representation of *f* for the unknown parameter β_i, which results in a linear model (Wood, 2006). One of the main differences to other models is that the linear predictor now has smooth functions for one or potentially all covariates/predictors (M. Clark, 2016, June 26). At the core of any GAM, there are basis functions, and smoothing splines, of which there are too many to go into detail here (e.g. cubic splines, and regression splines), to determine the 'wiggliness' of non-linear curves with a smoothing parameter lambda (λ) between zero and one, which had to be selected manually by the researcher (T. J. Hastie & Tibshirani, 1986), but can now be selected automatically. A λ of 0.6 often seems to yield good results (Larsen, 2015, July 30; Wood, 2006). With λ, we can control the balance between 'wiggliness' of the function *f* and goodness of fit to the data. As larger values of λ make *f* smoother, the curves also tend to have more bias; i.e. in-sample-error (T. J. Hastie & Tibshirani, 1986, 1990; Larsen, 2015, July 30).

Thin plate regression splines (TPRS, Wood, 2003) in particular, often yield excellent results with the mean squared error in mind by combining lower-level functions (e.g. linear, quadratic, cubic, or logarithmic) to fit a smoothed function (Winter & Wieling, 2016). This is probably why TPRS is the default setting in the **`mgcv`**-package to conduct GAMs in R (Wood, 2006). As for the interpretation of GAMs, each smooth (non-parametric variable) has a *p*-value to test whether it is significantly different from 0 and the 'edf' (effective degrees of freedom, which indicate the level of non-linearity. An edf of 1 indicates a linear pattern, and any edf > 1 indicates a non-linear pattern. Since there are no coefficients for the predictors, visualization of the model fit is crucial (Winter & Wieling, 2016). The interpretation of the GAM-output, as well as other important issues such as concurvity, selection of smoothing parameters, and model/variable selection by way of cross-validation will be addressed further in the results and analysis section. Naturally, the present work can only include a cursory introduction, and there is a lot more to be said about GAMs. Larsen (2015, July 30) and Clark (2016, June 26) offer two concise, yet thorough introductions to GAMs with examples in R, which should be complemented by Hastie and Tibshirani (1986, 1990) and Wood (2006), especially for output interpretation and more detailed explanations of core concepts.

3.10 Software

3.10.1 R/RStudio

All statistical analyses were run in the statistical environment R, version 3.4.1 "Single Candle," (R Development Core Team, 2017) with RStudio, version 1.0.153, (RStudio Team, 2017) as integrated environment. R has become the gold-standard for statistical computing, which is attributable not only to its open-source nature, but more importantly, to the fact that most new statistical methods are often first implemented in R and published as free packages (Tippmann, 2015). Beyond being very flexible and up-to-date in terms of statistical procedures, R also lets the user produce publication-quality graphs and animated output for easy and flexible data visualization.

The R-package, **`rmarkdown`** (Allaire et al., 2017), was used to keep track of every step of the analyses to be able to reproduce, and if necessary, amend any step along the way. Other packages used for exploratory data analysis and statistical analyses will be mentioned in the results and analysis section.

3.10.2 Language Processing

Twitter data collection and language processing was also implemented using the statistical programming language R by writing R-scripts that extract information as necessary. As mentioned above, these scripts essentially produced three language data sets: (1) text only, (2) one hashtag only, (3) and one data set with emojis. Linguistic Inquiry and Word Count (LIWC) was used to get the percentages of individual word categories. R was used because it is both accessible online and free and allows for the flexibility to write, change, and amend scripts to one's particular needs. To filter out English tweets, the **`cldr`**-package, version 1.1.0, (McCandless & Sanford, 2013a, 2013b), was used (also see above). To obtain measures such as Carroll's CTTR and Yule's K, R's **`koRpus`**-package, version 0.10-2, (Michalke, 2017a) was used in conjunction with **`TreeTagger`**, version 3.2, with the German parameter file encoded in UTF-8, (H. Schmid, 1994, 1995), a probabilistic POS tagger,[27] which uses Hidden Markov Models[28] and decision trees to tag words according to the part of speech. The `koRpus`-package functions as an R-wrapper, which accesses **`TreeTagger`**, whose output-objects can then be used further to obtain Carroll's CTTR and Yule's K, and many other measures of lexical diversity. While the **`koRpus`**-package has a built in tokenize-function, its range in terms of POS-tagging is very limited, which is why the use of the package together with the POS-tagger, **`TreeTagger`**, is highly recommended by the author (Michalke, 2017a) in order to obtain accurate results. **`TreeTagger`**'s English parameter file was trained on the Penn treebank

---

[27] Parts-of-speech-tagger.

[28] A Hidden Markov Model, or HMM, is a machine learning model that uses a Bayesian probability approach to, very simply put, determine the most likely sequence of words. It is thus closely related to Bayesian inference. Based on the original Markov Model, an HMM allows the user to study observed events (words from the input) and hidden events (such as POS-tags) much like Factor Analysis or Structural Equation Modeling. Because of this feature, they are considered crucial machine learning models in speech and language processing (Jurafsky & Martin, 2009).

tagset[29] (Penn Treebank Tag Set, 2017) while the German parameter file was trained on the Stuttgart-Tübingen tagset (German Tagsets, 2017; TreeTagger, 2017).

3.10.3 Version Control

R allows for full version control of individual scripts, which makes it possible to track any changes and go back to previous versions if necessary. Version control was implemented using Git and GitHub. Furthermore, scripts allow the user to add pseudocode in the form of comments preceded with a # to "label" each step of the analysis, which is indispensable for projects of this magnitude. This way, I was able to keep track of exactly what data were analyzed, what model was run, and where the output in the form of numbers and graphs came from, which is a lot more flexible and versatile than analyses run in SPSS, or any statistical program with a GUI.[30] This approach also supports the current paradigm of open science and quality control in the form of reproducibility. While other researchers will not be able to obtain my entire dataset, they will be able to reproduce my analysis with their own data sets.[31]

3.11 Tidy Data Structure

All data collection, data wrangling and cleaning, and analyses follow the tidy data paradigm (Wickham, 2014) — a now-current core concept in data science. That means data sets were organized in the following format: one row per observation/token, one column per variable. This paradigm ensures clean data sets conducive to data housekeeping, exploratory data analysis,

[29] A tagset is a list of parts-of-speech such as nouns, verbs, and prepositions. A tagset has a lot more word classes than what is usually used in syntactic analysis: The Penn Treebank has 45 parts of speech for instance. POS play a significant role in natural language processing because of the large amount of information they provide about the individual word and its neighbors (Jurafsky & Martin, 2009).
[30] Graphical user interface.
[31] The R-scripts are available upon request from the author.

and statistical analyses. Together with the tokens in Table 1, the two tables, below (Table 2 and Table 3) show how data were organized and maintained.

Table 2: *Participant Template Data-Frame for demographic/personality information*

part_id	gender	age	…	big_five	score	#_tweets	emoji	em_den	…	#_tweets	…
1	m	22	…	o	3.5	367	pos	24	…	367	…
1	m	22	…	a	1	367	pos	24	…	367	…
1	m	22	…	c	4	367	pos	24	…	367	…
1	m	22	…	n	2.5	367	pos	24	…	367	…
1	m	22	…	e	4.5	367	pos	24	…	367	…
…	…	…	…	…	…	…	…	…	…	…	…

…	time_on_twi	em_words	em_words_per	…	cttr	yules_k	…
…	1	pos	75	…	24	35.6	…
…	1	pos	75	…	24	35.6	…
…	1	pos	75	…	24	35.6	…
…	1	pos	75	…	24	35.6	…
…	1	pos	75	…	24	35.6	…
…	…	…	…	…	…	…	…

Table 2 illustrates how the data are structured in the data set. The participant's individual anonymous id-numbers are followed by demographic information. Then come the scores for the Big Five personality measure, which are followed by total number of tweets, emoji type, and density for example. This bridges the gap to word based measures from LIWC, such as type and number of emotion words. Finally, there are measures of Carroll's CTTR and Yule's K, as measures of lexical diversity. In Table 2, we only see all five levels of the big_five categorical variable. Naturally, other categorical variables, such as emoji, have two or more levels as well, which are left out for the sake of space as they would unnecessarily extend the table.

Table 3: *Tweet Template Data-Frame for the Entire Tweet-Data Set*

part_id	tweet_id	tweet	created	source	…
1	1	tweet text	2017-02-24	Twitter for iphone	…
1	2	…	2015-03-12	Instagram	…
1	3	…	2014-01-10	Twitter web cient	…
…	…	…	…	…	…
2	1	…	2017-01-01	Twitter for Android	…
2	2	…	2016-12-31	Instagram	…
2	3	…	2015-10-07	Twitter web client	…
…	…		…	…	…
3	1	…	2017-03-13	Twitter for iPhone	…

Table 3 contains every participant's tweets, including their date and the source of the tweet. In this context, it is important to mention that a tweet made on the Instagram app and then posted to Twitter is regarded just as much of an actual tweet as one made on Twitter's web client or Twitter's mobile apps, since what matters is what user-language ends up on Twitter, not particularly how. This procedure again ties in with the tidy data paradigm and the tidy text format according to Silge and Robinson (2017). This caveat notwithstanding, the source is an interesting variable used to determine where tweets predominantly come from. Additional information such as gender, age, and other demographic or other variables were added as needed in the analysis, see Chapter 4.

CHAPTER FOUR: FINDINGS AND DISCUSSION

4.1 Introduction

This chapter details the results of the study, including descriptive statistics, and the outcomes of the hypothesis tests, including inferential statistics. It begins with a description of the participant sample, demographic information on the participants (e.g. gender and age/age groups), and the Twitter data (e.g. total numbers, tweet-times, tweet sources). I then move on to the hypothesis tests, which are divided into four sections according to their grouping in the methodology chapter (Chapter 3): effects of personality on LIWC categories, gender effects and Twitter measures, gender effects and LIWC categories, and gender effects and word-based measures.

4.2 Participant Demographics

Originally, 202 participants submitted the Qualtrics survey. Of these 202 surveys, 161 were complete, with 75 data sets having valid Twitter accounts (this amounts to participant attrition of 62.9%).Many participants filled out the questionnaire including a Twitter-handle. However, after hand-checking the accounts, it turned out that some had either filled in a made-up user name, a user name that was not theirs, e.g. @therealdonaldtrump, or simply put in *Nein* 'no.' Thus it seems that participants did not have any issues filling out the questionnaire including the personality profile, but ultimately did not want that to be associated with their online presence on Twitter. This is reflected in research, which has shown that German social media users (72.4%) put a lot of emphasis on data protection and privacy in the social media sphere (Burda Forward, 2015). Filtering out the participants who tweeted predominantly in English brought the number down to 70 participants. Applying the exclusion criterion (minimum of 50 unique words) resulted in the loss of an additional seven participants. After an initial look at the tweets, one participant turned out to be an online marketer, who almost exclusively used a tweet-bot to create tweets. This participant was also excluded., bringing the total number of participant down to $N = 62$.

In the following, I present key descriptive statistics pertaining to the participants in this study, which provide important demographics of German Twitter users. As outlined in Chapter 3, the demographic variables collected via the questionnaire encompass gender, age, zip code, zip code2 (optional: the place where participants grew up if different from zip code), relationship status, citizenship, native language, education1 (school type), education2 (academic training/vocational training), and employment.

4.2.1 Participant Gender and Age

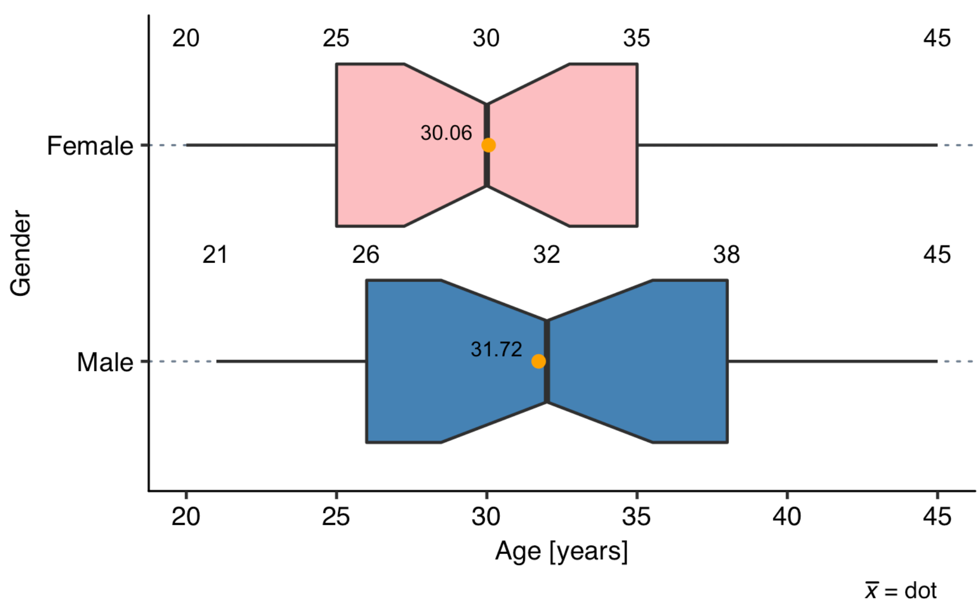

Figure 5. Participants by Gender and Age.

The dataset consists of 29 (46.8%) male and 33 (53.2%) female participants with a group difference of only four participants. The gender of the participants pertains to the socially enacted construct of gender as indexed by the participants' answers in the questionnaire. No conclusions about the participants' biological sex were drawn here. The age distribution in Figure 5 confirms Pfeiffer's (2009 as cited in Friedrichs, 2009) findings that German Twitter users were 32 years on average; female users, $\bar{x} = 30.06$, $\sigma = 6.92$, $Mdn = 30$;[32] male users, $\bar{x} = 31.72$, $\sigma = 6.45$, $Mdn = 32$, with an overall mean of 30.84 years, an overall standard deviation of 6.7 years, and an overall median of 30 years. An unpaired Welch two-sample t-test was run to determine if the group differences for age are statistically significant: they are not, $t(59.77) = 0.98$, $p = .33$, $d = .25$.

We can further infer from the boxplot (Figure 5) that the age distribution among participants is more symmetrical for the males than for the females (their top whisker is a little longer then the bottom whisker) with the female participants being a little younger than the male participants (cf. interquartile range); overall, age turned out to be nearly normally distributed.

One of the reasons German Twitter users are roughly 30 years on average might be that many of them were in their early-mid-twenties when Twitter first started in 2006 over ten years ago. As Twitter has arguably fallen out of favor with many younger users, especially 14–19-year olds, (Bauer, 2017, June 13), ranking behind Instagram (roughly 11 million users expected in 2017 in Germany (Business Insider, 2017, May 11)) and Snapchat in Germany, Twitter now seems to be 'reserved' for young, adult users in their mid–late twenties, and early–mid thirties.

---

[32] $\bar{x}$ = Sample mean, σ = Standard deviation, Mdn = Median

4.2.2 Participants' Relationship Status

Table 4: *Participants' Relationship Status*

Relationship status	Male *f* (%)	Female *f* (%)	Overall *f* (%)
In a relationship	14 (48.3)	16 (48.5)	30 (48.4)
Single	12 (41.4)	13 (39.4)	25 (40.3)
Married	3 (10.3)	3 (9.1)	6 (9.7)
Divorced	n/a	1 (3.0)	1 (1.6)
Overall	29 (46.8)	33 (53.2)	62 (100)

Note. Since some of the cells have small values, a Fisher's exact test was run to examine the relation between gender and relationship status ("In a relationship" and "Single"); it was not significant, $p = 1$, Cramer's $V = 0.013$, showing that the number of participants, who were either in a relationship or single did not differ significantly by gender.

Table 4 is interesting in that it illustrates that almost 50% of participants (48.3% of males and 48.5% females) are in a relationship compared to the second largest group, singles, with a little over 40% (41.4% of males vs. 39.4% of females). The numbers for married and divorced participants (male and female) seem almost negligible in comparison. Since social media are deeply embedded in our daily lives, it is not too surprising that many of the participants are either in a relationship or single. I venture that there is simply not enough data for married participants in this study, but it is conceivable that social media use declines when people are married due to other commitments/interests. Interestingly, research has found a correlation between the use of social media and increased divorce rates (Valenzuela, Halperna, & Katz, 2014). Since many divorces happen later in life (in 2015, the average age for women in Germany was 43.3 years and for men 46.3 years (Statista, 2017)), one possible explanation for the low number of divorced participants is simply that not enough of them fell in the age bracket selected for the study (18–45), and that younger users are less likely to be married yet. An increase in the use of Twitter post-divorce is expected to a lesser extent than other social media sites since it does not serve the same socialization purposes that Facebook, Snapchat, Instagram, and especially online dating platforms, such as Tinder, afford in this context.

4.2.3 Participants' Geographical Distribution across Germany

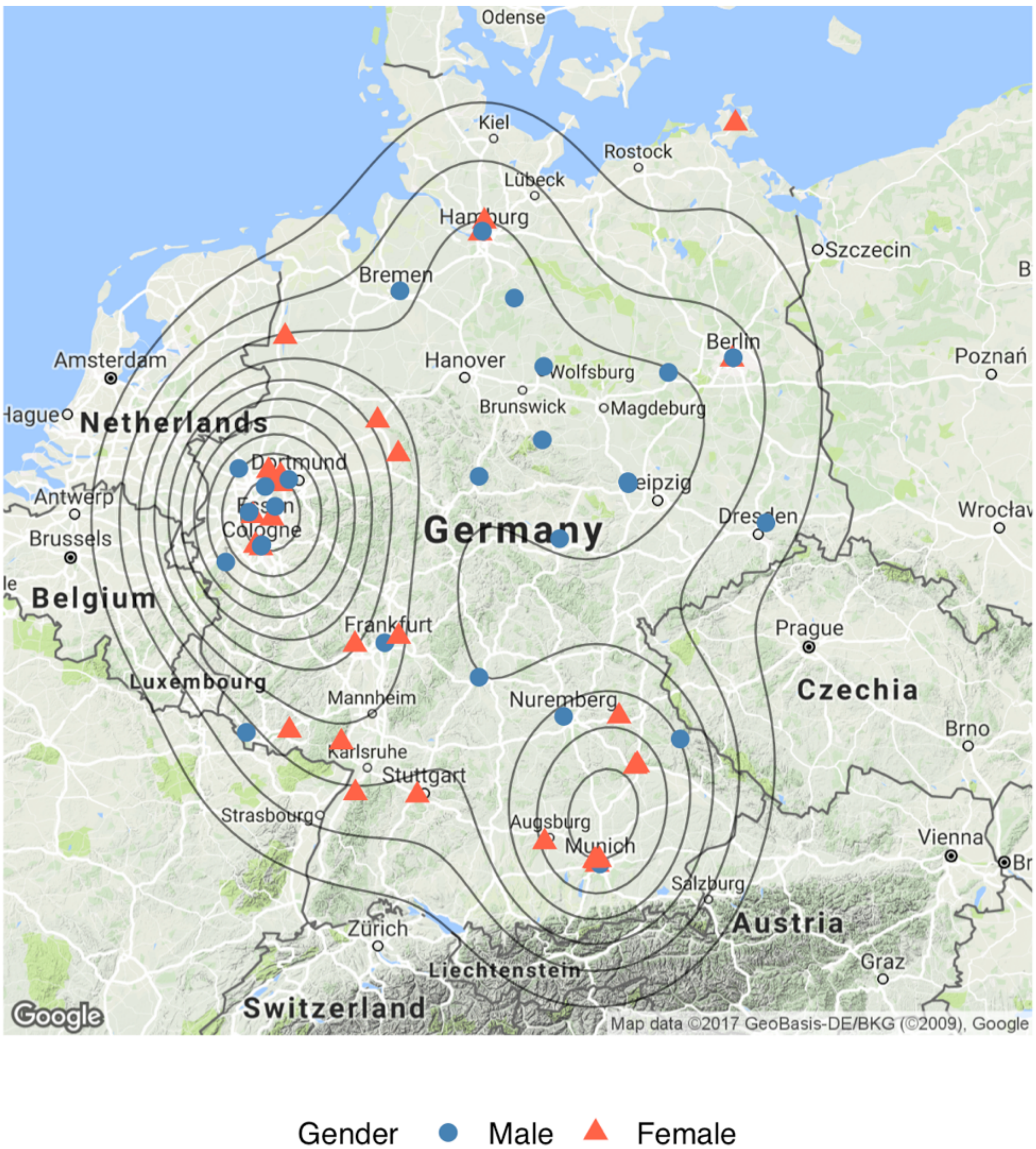

Figure 6. Participant Distribution across Germany with Densities.

All participants, $N = 62$, are Twitter users, who come from and live in Germany and whose native language is German. As can be inferred from Figure 6, there is a definitive trend of Twitter users clustering around larger metropolitan areas (cf. density lines), especially in the northwest (Cologne, Essen, and Dortmund), which has the highest density of users, followed by southern Germany (i.e. Bavaria: Munich, Nuremberg, Regensburg, and Augsburg) although, here, participants are more dispersed across a wider geographical area. These bigger clusters are trailed by smaller clusters; e.g. southwest/center of Germany (Frankfurt, Karlsruhe, Stuttgart), Hamburg in the north, and Berlin in the northeast. Except for an area extending from the center to the North of Germany, where data come predominantly from male users, data are available from both male and female users in the discernible clusters. Twitter data are available for almost all 15 states with differing frequencies except for Schleswig-Holstein with its state capital Kiel in the north.

In addition, Figure 7 below shows where participants moved if they moved away from the place they grew up in (measured by the two different zip codes, which were then matched with their respective longitudes and latitudes). In the context at hand, this has potential social implications: (1) most participants stayed put, mostly clustering around larger metropolitan areas, and (2) the participants in this study who did move were predominantly drawn to the aforementioned large metropolitan areas, suggesting that Twitter in Germany is predominantly used by urbanites (even though not all of them are urbanites 'by birth'). This aligns nicely with the density plot in Figure 6 above, in which most participants cluster around big cities. For more in-depth variationist analyses of linguistic patterns, a visualization like Figure 7 is crucial to explain and detect dialectal 'outliers' when investigating regional variation in tweets; the participants who moved from Berlin to Frankfurt or form Leipzig (Saxony) to Munich (Bavaria) moved into a different dialectal area, but arguably retained some, if not all of their own dialectal features, i.e. pronunciation, lexicon, and syntactic patterns, among others.[33]

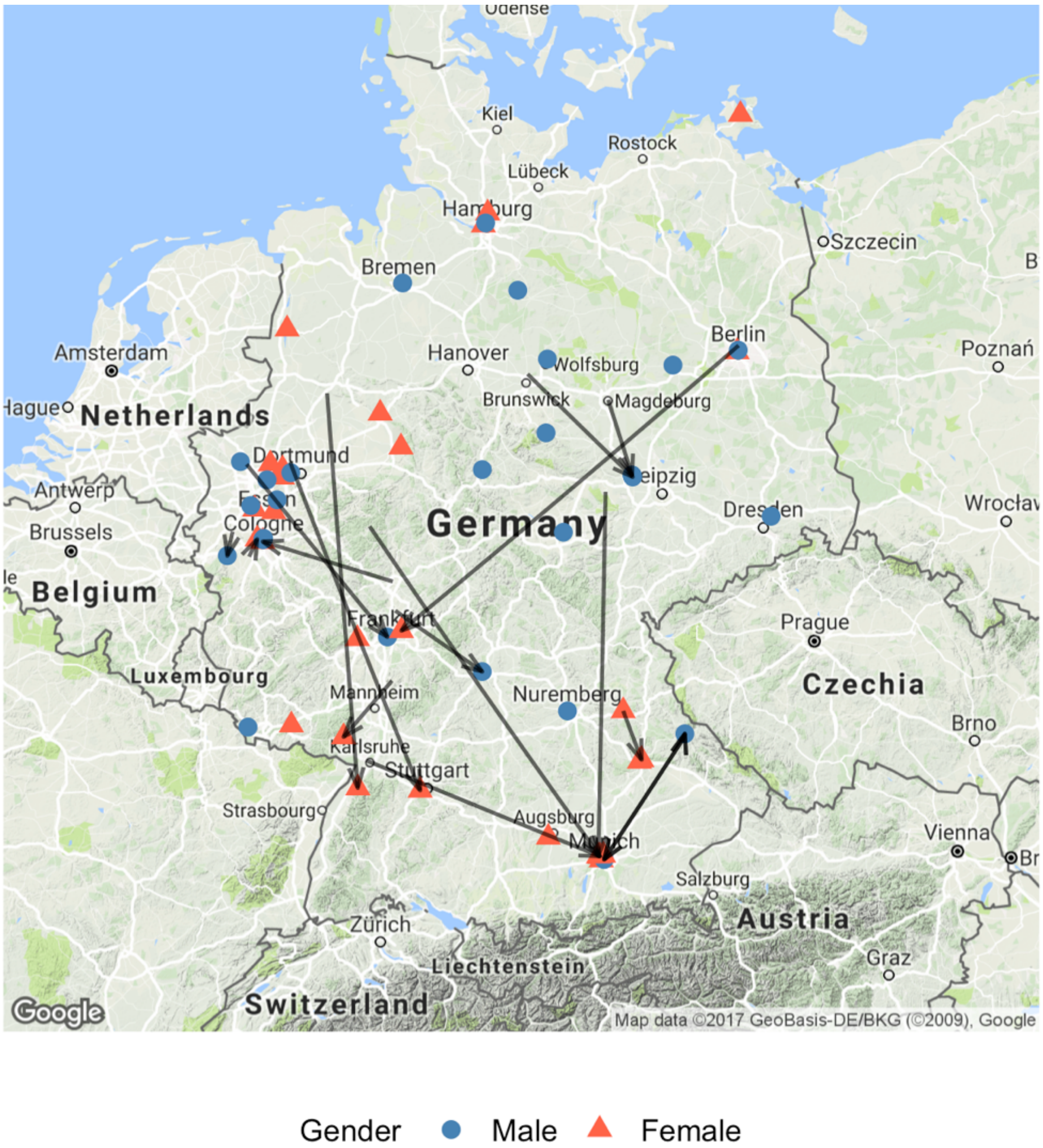

Figure 7. Participants' Origins and Residences (Default Locus of Tweets).

[33] This, of course, usually depends on the number of people moving from one place to another. Mass influx of outsiders to a certain dialectal area can trigger the decline of a regional dialect (or certain features thereof), e.g. in the Outer Banks of North Carolina, or Raleigh, NC (Dodsworth, 2013; Wolfram & Schilling-Estes, 1997, 2006).

4.2.4 Participants' Education and Employment

Table 5: *Participants' Education and Employment by Gender*

Education	Male *f* (%)	Female *f* (%)	Overall *f* (%)
Education 1			
High school (Abitur/FOS/BOS)	22 (75.9)	30 (90.9)	52 (83.9)
Mid-tier – secondary (Realschule)	5 (17.2)	3 (9.1)	8 (12.9)
Lowest tier (Hauptschule)	2 (6.9)	n/a	2 (3.2)
Education 2			
University	10 (34.5)	20 (60.6)	30 (48.4)
University of applied sciences	5 (17.2)	2 (6.1)	7 (11.3)
Apprenticeship	7 (24.1)	5 (15.1)	12 (19.3)
No degree (incl. students)	7 (24.2)	6 (18.2)	13 (21)
Employment			
Full time (40 hrs.)	21 (72.4)	18 (54.5)	39 (62.9)
Part time (20 hrs.)	2 (6.9)	2 (6.1)	4 (6.5)
Student	6 (20.7)	13 (39.4)	19 (30.6)
Overall	29 (46.8)	33 (53.2)	62 (100)

Note. Separate Fisher's exact tests were run on gender and education1, education2, and employment. All tests came back statistically non-significant, so we fail to reject H_0 (the relative proportions of gender are independent from each of the other variables, education1, education2, and employment): p = .23, Cramer's V = .26; p = .19, Cramer's V = .28; p = .28, Cramer's V = .20.

Table 5 reveals some interesting aspects about the participants: (1) Well over 80% of the participants graduated from an academic high school, making them eligible to pursue university studies. (2) Between genders, ~91% of female participants obtained high school diplomas, trailed by the males at ~76%. The other types of education are very low when contrasted with those numbers. This trend bleeds over into continued education, where ~67% of females have a university education (university and university of applied sciences) or are in the process of obtaining one (~27%), or are in the workforce (~40%), ~52% of males have done/are doing the same (~32% are graduates or are working, and ~20% are students). (3) It is thus not surprising that, when it comes to employment, the females lead the group of students, with ~39% of female participants currently working towards a degree or possibly an advanced degree while only ~21% of males following that same path. (4) Overall, well over half (~60%) of all participants have or are obtaining a university education and ~63% of all participants are employed full time. Table 5 is thus also a reflection of recent and not so recent developments in the educational population in Germany, and many other countries in the western world for that matter, in which more females than males graduate from high school (WDR, 2016) and, consequently, also from universities. The situation seems to be reversed in the lowest-tier of education (Hauptschule), from which more males graduate (WDR, 2016). This is also mirrored in Table 5, if only to a small extent. At this point, I propose to relabel Eurostat's (2016) terminology (high school education = less educated) to fit the German context with a three-tiered school system, in which the academic high school is the eligibility requirement for university admission. This makes its graduates more educated than potential international counterparts, where the three tiers are conflated into one, as is the case in the United States, for example. For a short overview about the diverse ways paths in the education system, see Figure 8 below.

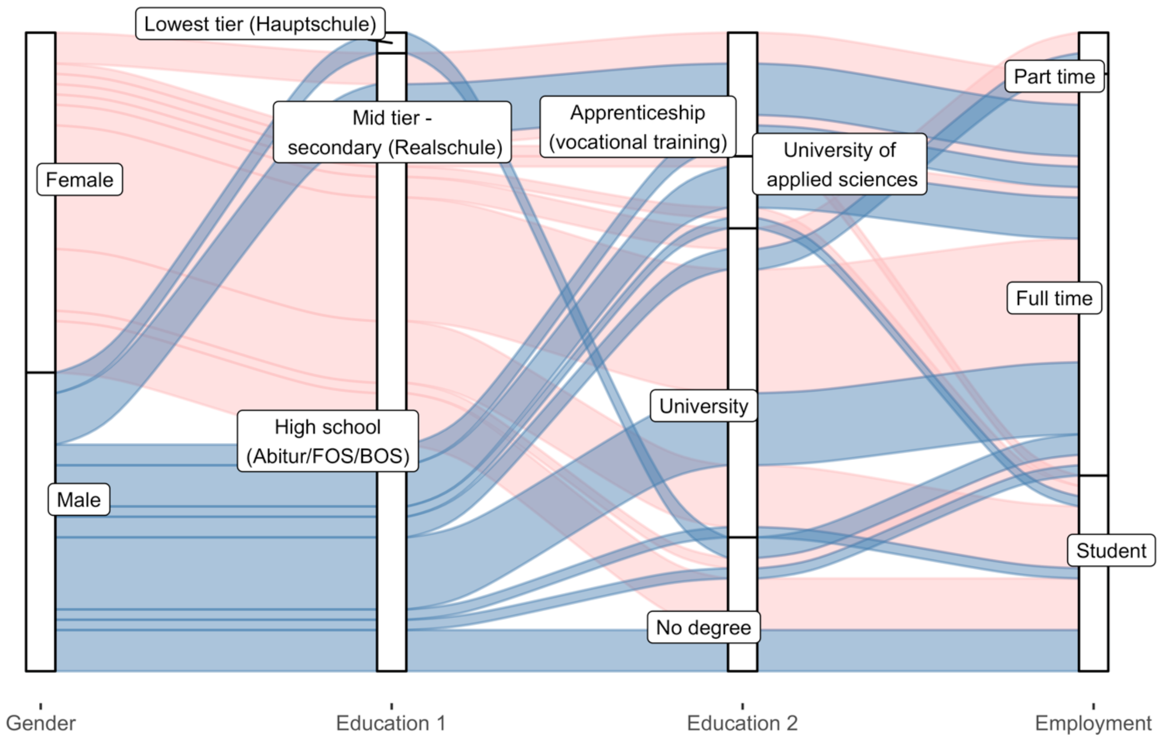

Figure 8. Alluvial Diagram: Participants' Educational Paths to Employment by Gender.

Figure 8 also shows the path of the two male participants who attended school in the lowest tier (Hauptschule), did not obtain a degree, and went into the workforce right away. They represent the least educated participants in the study. It should be mentioned that the 'Student' group encompasses participants who are still students (at the undergrad level) and/or grad students. Similarly, the 'Full time' and 'Part time' groups are also comprised of students, at least in part. Against a German backdrop, the numbers in Table 5 seem to confirm Eurostat's (2016) and Pfeiffer's (as cited in Friedrichs, 2009; Hilker & Raake, 2010) findings, as most of the participants are well-educated (males and females) Twitter users.[34]

[34] Since every survey is susceptible to certain biases, it is conceivable that simply more well-educated participated, or that a social class difference results in educational differences.

4.3 Participants' Twitter Habits

In the following, I present some important Twitter metrics for the participants: e.g. how long they have been on Twitter, how often they check Twitter, what they think the function of a hashtag is, and how much time they spend on Twitter per day in minutes.

Table 6: *Participants' Twitter Metrics*

Twitter metric	Male *f* (%)	Female *f* (%)	Overall *f* (%)
Years on Twitter			
Less than a year	5 (17.2)	8 (24.2)	13 (21)
1-2 years	2 (6.9)	4 (12.1)	6 (9.7)
2-3 years	2 (6.9)	5 (15.2)	7 (11.3)
More than 3 years	20 (69)	16 (48.5)	36 (58.1)
Check Twitter			
< once/day	2 (6.9)	8 (24.2)	10 (16.1)
Once/day	11 (37.9)	10 (30.3)	21 (33.9)
Several times/day	16 (55.2)	15 (45.5)	31 (50.0)
Function hashtag			
Tag	24 (82.8)	22 (66.7)	46 (74.2)
Comment	n/a	1 (3)	1 (1.6)
Both	5 (17.2)	10 (30.3)	15 (24.2)
Overall	29 (46.8)	33 (53.2)	62 (100)

Note. Separate Fisher's exact tests were run on gender and years on twitter, check twitter, and function hashtag. All test came back statistically non-significant; thus, we fail to reject H_0 (the relative proportions of gender are independent from the each of the other variables, years on twitter, check twitter, and function hashtag): $p = .42$, Cramer's $V = .21$; $p = .35$, Cramer's $V = .24$; $p = .24$, Cramer's $V = .20$.

Table 6 reveals a couple of interesting things: (1) While the clear majority of all participants (58.1%) have spent more than three years on Twitter overall, the situation is a little more diverse when looking at gender. Here, the males lead the way with 69% on Twitter more than three years, and the second largest group (17.2%) on Twitter less than a year. Of the female participants, almost 49% have spent more than three years on Twitter. However, the second largest group (24.4%), who have spent less than a year on Twitter and the other two groups, 1–2 years and 2–3 years, respectively, still make up about 27% combined. The males in the latter two groups only make up roughly 14%. As the Fisher's exact tests have revealed (see Table 6, Note), the differences are not large enough to be significant. This can potentially be attributed to either the specific participant sample, or conceivably, the possibility that the females are lagging behind in adopting/utilizing Twitter as a social medium, both increasing the number of participants in the 'less than a year' group. (2) While half of all participants check Twitter several times per day (50%), the situation is a little different for the males and females again. Among the male participants, the majority check their Twitter accounts several times a day (55.2%). Here, the highest value for the females (45.5%) can be misleading, as it is the largest percentage. However, the two other groups in this category combined (< once/day and once/day) make up more than 50% of the female participants, meaning that most of the female participants only checked Twitter once per day or even less than that. (3) Regarding what the German Twitter users in this study think hashtags are used for, almost three quarters (74.2%) of all participants indicated they are used to tag tweets, while the remaining quarter believe they are used as both tags and commentary tags. Restricting the focus to participant gender, there are some slight differences (not statistically significant, see Table 6, Note). Roughly 82% of males think hashtags are being used to tag tweets compared to roughly 17% who think they are used for both.

With the females, those two percentages are more divergent: roughly 66% think they are used for tags, while roughly 30% think they are used for both tagging and commenting. This might be attributable to female Twitter users being better versed when it comes to social media use, while, at the same time, being more attune with and caring more about the intricacies and unwritten conventions of hashtag usage. These numbers seem to confirm Shapp's (2014) and Herring and Paolillo's (2006) findings that males use more tag-hashtags in more informational language. However, this is only based on the perceptions participants have about the use of hashtags. See hypothesis testing below for further details on this issue.

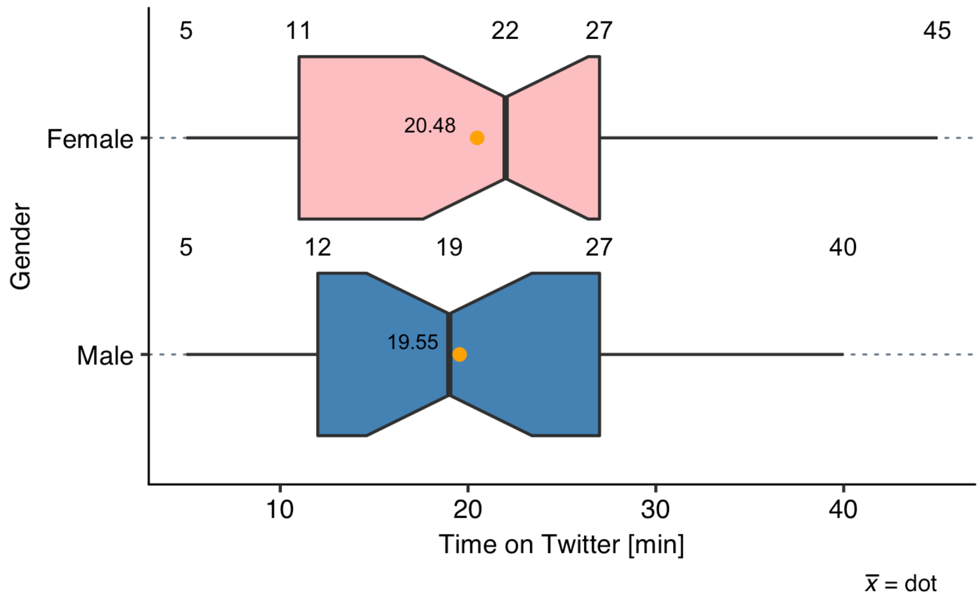

Figure 9. Amount of Time [min] Spent on Twitter per Day.

Figure 9 illustrates that, while there are some divergent findings for how often males and females check their Twitter accounts, they spend virtually the same amount of time on Twitter when they check their accounts (see interquartile range), with the upper quartile for the females having a higher maximum (45 min) compared to the males (40 min). Correspondingly, a Welch two-sample t-test did not produce statistically significant results, $t(59.89) = -0.36$, $p = .72$, $d = -.10$. Looking at the means, there is only a small discernible difference between males ($\bar{x} = 19.55$, $\sigma = 9.74$, $Mdn = 19$) and females ($\bar{x} = 20.48$, $\sigma = 10.65$, $Mdn = 22$), with the standard deviation being a little higher for the females by virtue of the higher variance in the time they spend on Twitter. The overall mean is 20.05 minutes, with an overall standard deviation of 10.16 minutes. The finding that German males and females spend about the same amount of time on Twitter agrees with the results of other researchers (Burda Forward, 2015; Casey, 2017) for German and US Twitter users.

4.4 Overview of Participants' Personality Scores

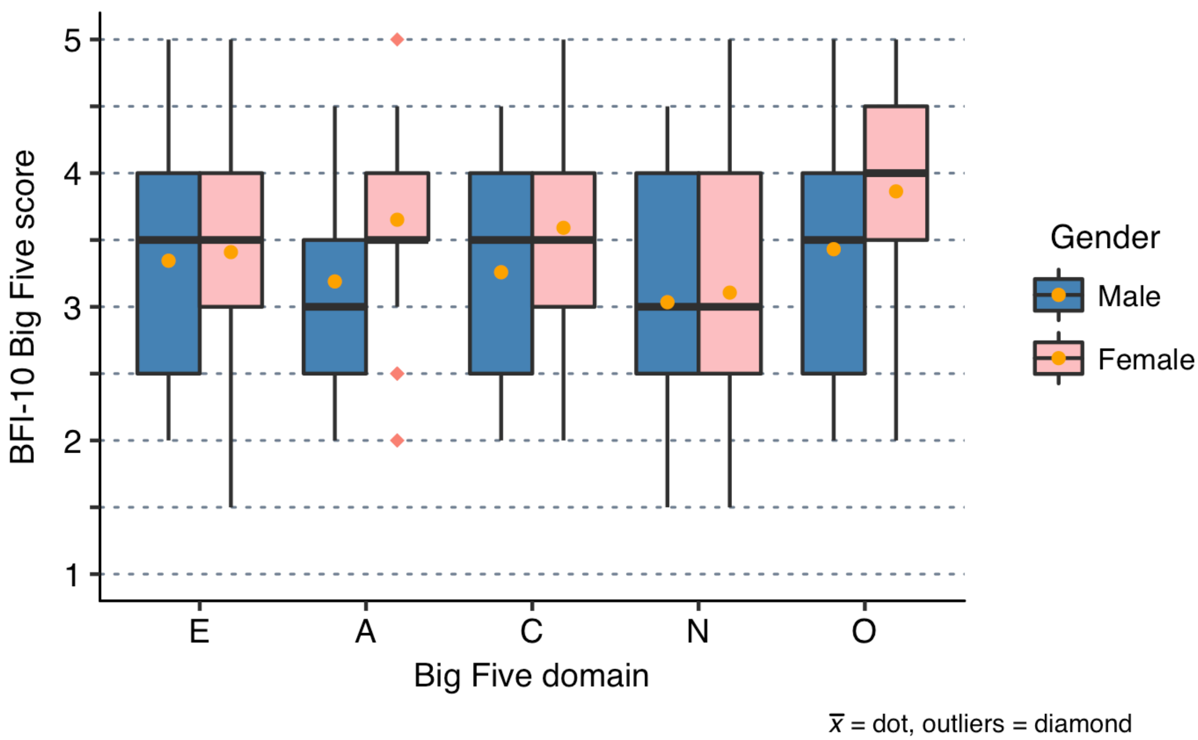

Figure 10. Participants' BFI-10 Big Five Scores.

To reiterate, the Big Five factors comprise extraversion (E), agreeableness (A), conscientiousness (C), neuroticism (N), and openness (O). These factors are usually measured with Likert-scales, ranging from one (strongly disagree) to five (strongly agree). Participant answers for each factor are averaged to obtain individual scores for each factor. Jointly, the resulting five scores represent an individual's personality (John & Srivastava, 2001). A high score on extraversion means that an individual is sociable, outgoing, talkative, and assertive, while a high score on agreeableness indicates that an individual is cooperative, helpful, and nurturing. People scoring high on conscientiousness are responsible, organized, hard-working, and reliable. Individuals, who have high scores on the neuroticism trait are anxious, insecure, and sensitive, while people, who score high on openness are curious, intelligent, and imaginative (Golbeck, Robles, & Turner, 2011).

Figure 10 reveals some striking insights into the distribution of personality factors and BFI-10-scores for the gender variable. Extraversion (E) has an overall mean of 3.38, an overall standard deviation of 0.91, and an overall median of 3.5. While the interquartile range for the males is a little greater (2.5–4) than for the females (3–4), the female group has participants with scores that are as low as 1.5 for the BFI10 score. However, their means and standard deviations are very close, suggesting very little difference (males: $\bar{x} = 3.35$, $\sigma = 0.95$, *Mdn* = 3.5; females: $\bar{x} = 3.41$, $\sigma = 0.89$, *Mdn* =3.5). Existing research has shown that differences in extraversion between males and females do exist, however small, with women scoring somewhat higher than men (Costa & McCrae, 1992; Costa, Terracciano, & McCrae, 2001; Weisberg et al., 2011). The same picture presents itself for conscientiousness (C), which has an overall mean of 3.44, an overall standard deviation of 0.75, and an overall median of 3.5. While the mean differs slightly between males and females (3.26 vs. 3.59), their standard deviations are virtually the same, suggesting a relatively homogenous spread of the data (0.73 vs. 0.74) with the same median (3.5). These findings are reflected in existing research (women score somewhat higher than men), although they are not consistent across all cultures (meta-analysis in Costa et al., 2001; Feingold, 1994). Another personality factor where both genders yielded the same median (3.0) is neuroticism (N). Here, the overall mean is 3.07, the overall standard deviation is 0.84, and, accordingly, the overall median is 3.0. Again, there are slight differences between males and females, suggesting that the females are a little more neurotic than the male participants (males: $\bar{x}$ = 3.03, σ = 0.81, *Mdn* = 3.0; females: $\bar{x}$ = 3.11, σ = 0.88, *Mdn* =3.0). This is also mirrored in existing research (Costa et al., 2001; Feingold, 1994).

The biggest differences between males and females was agreeableness (A, overall $\bar{x}$ = 3.44, overall σ = 0.77, overall *Mdn* = 3.5) and openness (O, overall $\bar{x}$ = 3.66, overall σ = 0.84, overall *Mdn* = 4.0). For agreeableness, the mean for males was 3.19 (σ = 0.71, *Mdn* = 3.0), which is lower than that of female participants ($\bar{x}$ = 3.65, σ = 0.77, *Mdn* = 3.5) by 0.46 points on the BFI10-score. This suggests that the female participants in this study are, on average, more agreeable than the male participants. Figure 10 shows that the interquartile range for the females essentially starts at the top end of the male interquartile range, with the median being its lowest point. In the past, women have been found to score higher on agreeableness consistently (Costa et al., 2001; Feingold, 1994). For example, Costa et al. (2001) ran a meta-analysis of studies on the Big Five and gender differences spanning 26 cultures (*N* = 23,031), including both college-age and adult participants. They showed that women were consistently higher in agreeableness. For openness, the situation is similar, albeit, here, the interquartile ranges overlap to some extent. The mean for males was 3.43 (σ = 0.87, *Mdn* = 3.5), which is, again, lower than that of female participants ($\bar{x}$ = 3.86, σ = 0.77, *Mdn* = 4.0) by 0.43 points on the BFI10-score, suggesting that the female participants were, on average, more open than their male counterparts (this also shows in the interquartile range for both groups even though they have values as high as 5 and as low as 2). No significant gender differences were found for the openness domain in previous research (Costa et al., 2001; Feingold, 1994).

An analysis of variance (ANOVA) was run to test whether the intuitions gleaned from Figure 10 translate into statistically significant findings.

Table 7: *ANOVA Results for the Effect of Gender and Personality Factor on BFI-10 Scores*

Variable	*df*	*F*	ω^2	*p*
Gender	1	8.63	.03	$\leq .01$
Personality factor	4	4.16	.04	$\leq .01$

Table 7 shows that there was a significant main effect of gender on BFI10-scores. A Tukey HSD *post hoc* test confirmed that finding, $p \leq .01$. In addition, there was a significant main effect of personality factor on BFI10-scores; however, the Tukey HSD *post hoc* test did not reveal any significant differences between individual personality factor combinations and BFI10-scores. While the ANOVA revealed that there are statistically significant differences between gender and BFI10-scores as well as personality factor and BFI10-scores, the omega squared effect size values indicate that the difference is not very big, confirming the story Figure 10 is telling, which is also mirrored in the mean values, and existing research (Costa et al., 2001; Feingold, 1994). In the context of this study, this means that the suspected gender interactions pertaining to personality scores are in fact happening.

4.5 Summary Participants

Overall, the participants in this study follow a nearly normal age distribution with an almost even gender split (46.8% males vs. 53.2% females). The participants are predominantly in a relationship (48.3%) or single (40.3%) with the rest being married or divorced. Good geographic coverage of Germany (major metropolitan areas) was accomplished (Figure 6) with participants moving to larger cities, if at all (Figure 7). The participants are well-educated with 83.9% having finished highs school and a further 59.7% having or getting university education and 62.9% being employed full time. In terms of Twitter usage, 58.1% have been on Twitter longer than three years and 50% checking Twitter several times a day. In addition, both males and females spend roughly the same amount of time on Twitter. Most participants thought that the hashtag was exclusively used for tagging purposes (74.2%). For the above measures, there were no significant gender differences. There was a small, but statistically significant gender difference for Big Five scores, indicating that, between male and female participants in this sample, there are, in fact, discernable interactions between gender and Big Fives scores going on.

4.6 Overview of the Tweet-Corpus

This section provides an overview of the tweet corpus, which was collected from participants' individual twitter accounts after Qualtrics data collection had been concluded.

Table 8: *Participants' Total Number of Tweets by Gender and Age Group*[35]

Age groups	Male *f* (%)	Female *f* (%)	Overall *f* (%)
20–24 years	1,568 (13.4)	489 (6.0)	2,057 (10.4)
25–35 years	4,255 (36.4)	5,947 (73.5)	10,202 (51.6)
36–45 years	5,844 (50.2)	1,658 (20.5)	7,513 (38.0)
Overall	11,678 (59.1)	8,094(40.9)	19,772 (100)

[35] Age groups were formed from the continuous age variable for visualization and table presentation purposes.

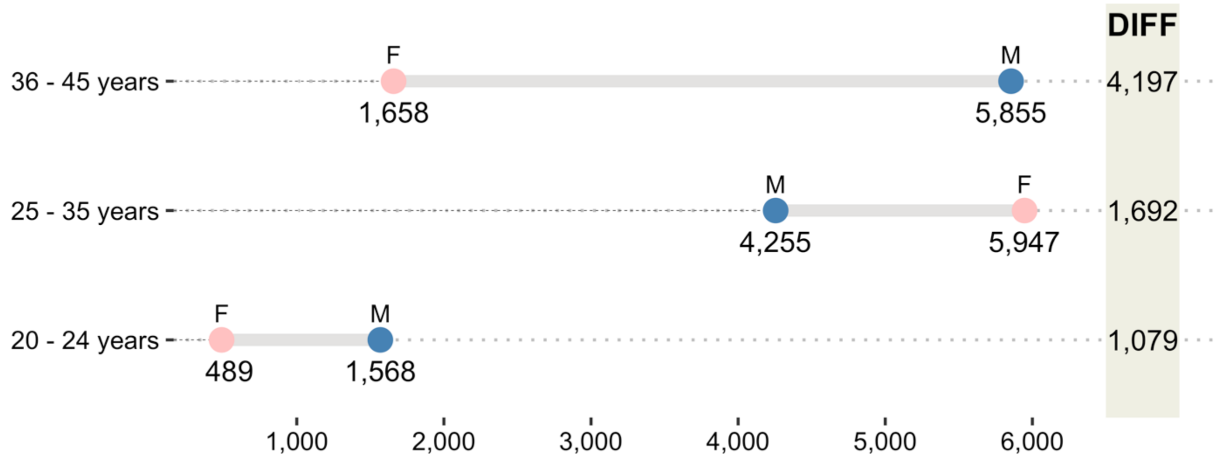

Figure 11. Differences in Number of Tweets by Gender and Age Group.

All in all, the tweet-corpus comprises 19,772 tweets. The male participants produced more (~59%) tweets than the female participants (~41%), with an overall mean of 318.90, an overall standard deviation of 493.31, and an overall median of 118.5. A negative binomial regression model[36] revealed that age (continuous) significantly predicted number of tweets, b = 0.06, p < .001. Gender, however, turned out not to be a significant predictor, b = -0.48, p = .11. While gender was not significant as a predictor, the negative coefficient does reflect the lower number of tweets for female participants (males: $\bar{x}$ = 402.69, σ = 601.69, *Mdn* = 140; females: $\bar{x}$ = 245.27, σ = 367.64, *Mdn* = 87). Interestingly, the male participants in the 36–45-year age group produced the most tweets while, in the female group, the participants in the 25–35-year age group produced the clear majority of tweets. I venture that the interesting piece of information here is that it is younger females, who, according to their age group (25–35 years),

[36] Negative binomial models are routinely used for count data, which, by their very nature, are bound by zero, `glm.nb(tweet_num ~ age + gender, data = diss_data)`; A likelihood ratio test comparing it to a Poisson regression model with the same predictors showed that the negative binomial model was a better fit for the data (also accounting for overdispersion), $\chi^2(4) = 24735, p \leq .0001$. The AIC for the negative binomial model was lower than the AIC for the Poisson model.

are likely to be professionals (see Figures 13–15 below, and hypothesis tests on word use and LIWC categories, which reflect this as well) using Twitter, whereas the male participants using Twitter are older (36–45 years) than their female counterparts (in terms of number of tweets). However, males in the 25–35-year age group make up the second largest group among males with 36.4% of tweets with 'only' a difference of about 1,700 tweets. This led me to conclude that those minor differences are mostly attributable to outliers, and not a particular usage pattern among males and females in different age groups.

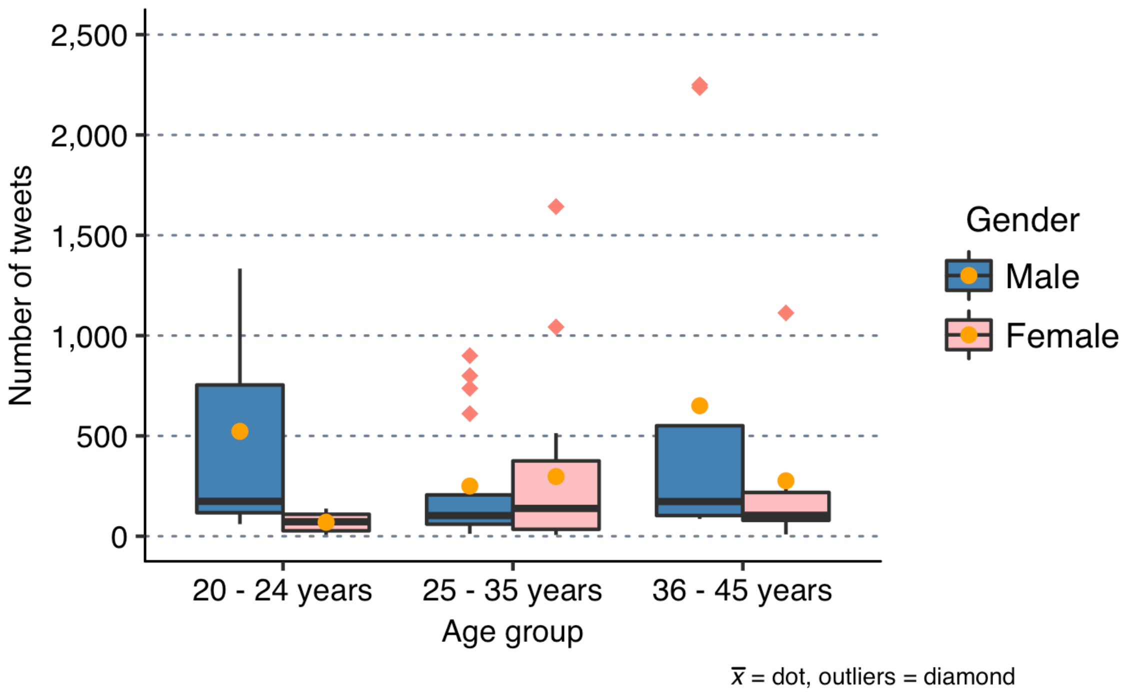

Figure 12. Participants' Overall Tweet Numbers by Age Group and Gender.

Figure 12 nicely illustrates how 'messy' the distribution of tweets really is. Interestingly, outliers are only present in the 25–35-year age group and the 36–45-year age group. The presence of outliers, however, bears testimony to the great amount of variation in the tweet data (also reflected in the means and standard deviations, see above). Considering this, it makes sense to look at the medians (overall = 118.5, male = 140, female = 87), rather than the means which are not only a lot closer together for the age groups and the genders, but also provide a measure that provides the true middle of the number of tweets and is not as susceptible to skew by outliers.

4.6.1 Participants' Tweets across Time

The following three figures, Figures 13–15, should be considered in tandem to get the best possible idea of when participants produced tweets and what possible implications this information has.

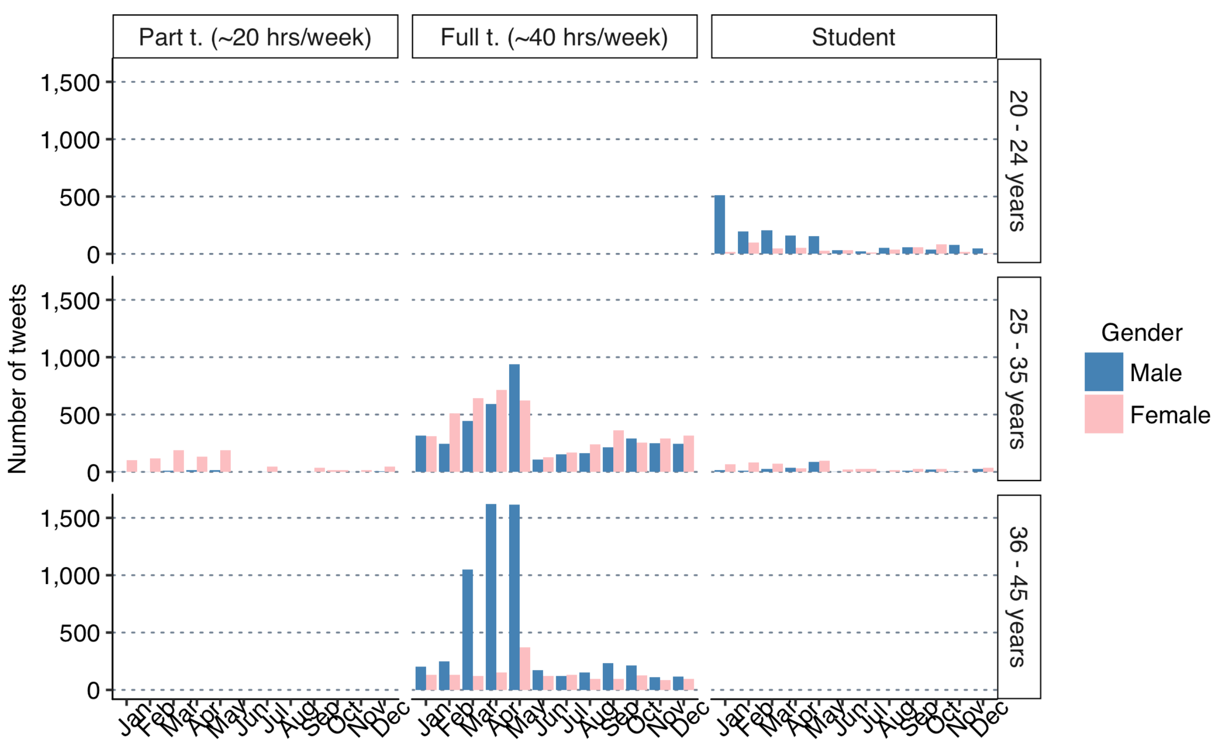

Figure 13. Participants' Tweets throughout the Year by Age Group and Gender.

When looking at Figure 13, we can make out a distinctively higher number of tweets during the colder months of the year across the board. This is most prominent for the 25–35-year olds and 36–45-year old participants (males and females) who are employed full time. (also cf. Table 8 above, which mirrors this in numbers). This could indicate pronounced leisure behavior displayed during the colder months of the year dropping drastically after the month of May for all participants and gender/age groups, which would essentially rule out consistent 'professional' use for promotion, self-promotion, and marketing throughout the year.

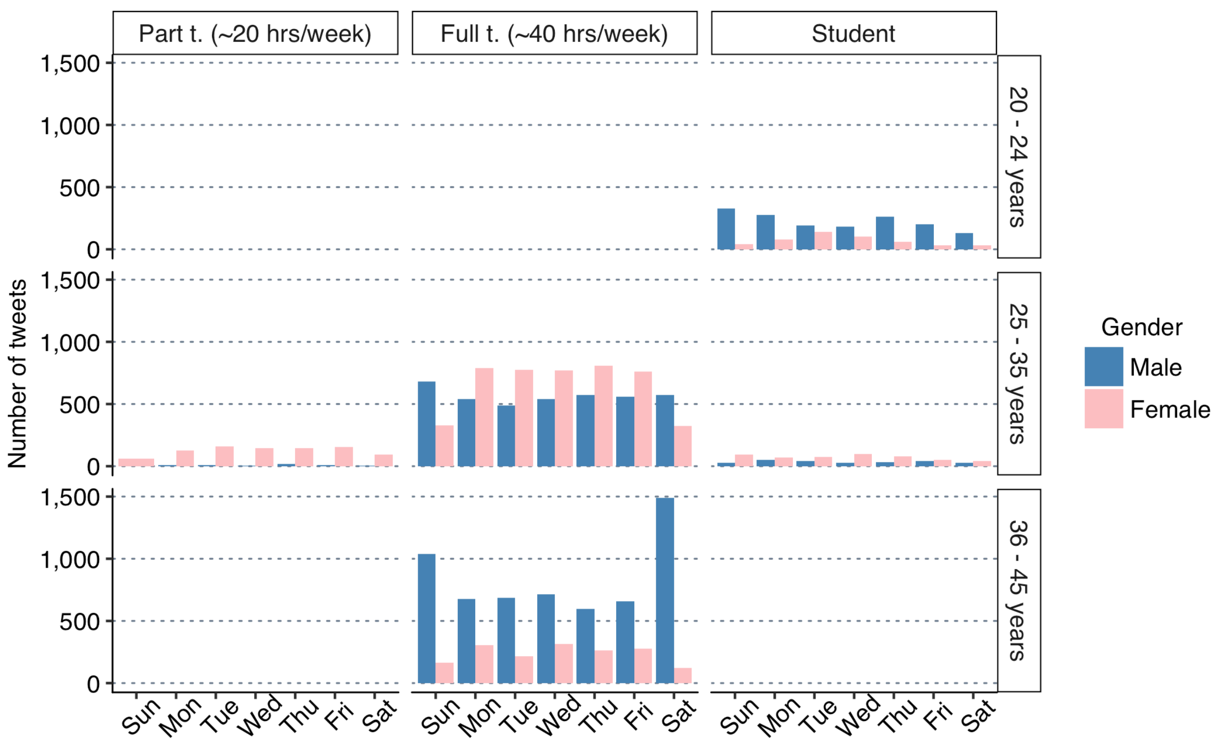

Figure 14. Participants' Tweets throughout the Week by Age Group and Gender.

When digging deeper and narrowing down the time frame, Figure 14 reveals that, factoring in the larger number of tweets from male participants, full time workers in the 25–35-year age group tweet relatively consistently throughout the week, while the males in the 36–45-

year age group dominate the picture especially on Sundays and Saturdays, which could again hint at a leisure behavior, instead of a more professional application of Twitter as a communication/media tool. The 20–24-year age group does not seem to show any interesting patterns other than a slight peak on Sundays for male participants and a slightly increased number of tweets for female students on Mondays, Tuesdays, and Wednesdays. Female student participants dominate the 25–35-year age group, if only slightly while the distinctly dominate the part time workers in the same age group with increased activity during the week (let us remember that the part time and full time work groups could potentially also encompass some (grad) students). This is also reflected in research: Zhu (2010, January 20)[37] found that most people tweet during the week with numbers going down over the weekend.

Finally, Figure 15 below provides insight into what hours of the day most tweets are produced and by whom. A negative binomial model[38] controlling for gender and age, confirmed that the hour of day is a significant predictor of number of tweets, $b = 0.05$, $p < .0001$. In addition, age turned out to be a significant predictor for number of tweets, $b = 0.05$, $p < .0001$, while gender, not surprisingly, is not a significant predictor, $b = -0.17$, $p = .08$ (also see negative binomial model above). Contrasting Figure 15 with Figure 14 above, it seems as if the females working full time in the 25–35-year age group tweet consistently during working hours, while the men in the same age group have a dip around noon and their nadir towards the evening hours.

[37] Hubspot analyzed five million Twitter accounts and six million tweet reports to come to their conclusion (Zhu, 2010, January 20). Although the report is over seven years old, the findings still seem to be valid.

[38] `glm.nb(n ~ hour + gender + age, data = t2)`; A likelihood ratio test comparing a Poisson regression model with the same predictors showed that the negative binomial model was a better fit for the data (also accounting for overdispersion), $\chi^2(1) = 29862, p \leq .0001$. The AIC for the negative binomial model was lower than the AIC for the Poisson model.

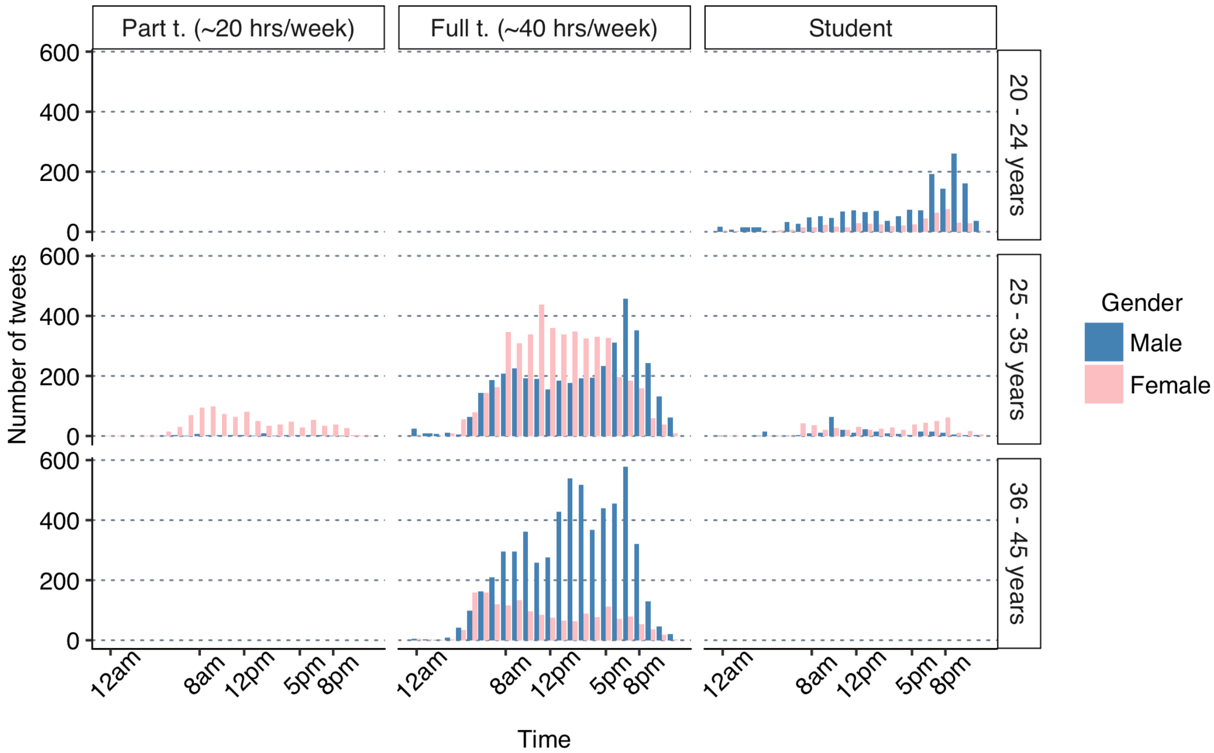

Figure 15. Participants' Tweets throughout the Day by Age Group and Gender.

Full time workers in the 36–45-year age group paint a different picture; the men dominate both the weekdays and especially the weekends with much higher numbers of tweets throughout the working hours. Part of that is certainly attributable to these men tweeting on the weekends. It appears that, when it comes to using Twitter as a professional/business tool, it is not so much a question of gender, but one of age with the females in the 25–35-year age group leading the way. Both the 25–35-year-old and the 36–45-year-old male and female participants seemingly follow the research on when the best times to tweet are in regards to receiving clicks on their tweets (K. Lee, 2016, April 27).[39] This comes with one caveat: a decreased number of tweets was sent during the late evening hours/night, which, according to Lee (2016, April 27) is the best time for

[39] The team at Buffer Social analyzed 4.8 million tweets and found the best time for most clicks to be the early morning hours, and for most retweets and favorites to be evenings and late at night (K. Lee, 2016, April 27).

getting retweets and likes. This is likely unbeknownst to the users, and not consistent as the numbers increase more after the early morning hours for the 25–35-year-old full time worker males display a distinctive dip during lunch hours picking up again after 12pm, while this is not true for the female participants. The 36–45-year-old age group does not mirror this 'lunch time' behavior, although the females do have a decrease in number of tweets starting around 9am and picking up again after around 1pm. This is also true for the female participants in the part time group (25–35 years). Diverging from this pattern are only the 25–35-year-old students (males and females), who display a slight dip during working hours, whereas 20–24-year-old students (both male and female) display a definitive increase of tweets after traditional working hours (starting around 5pm).

4.6.2 Participants' Twitter Measures Used in Hypothesis Testing

Table 9: *Participants' Twitter Measures for Analysis*

	Age groups						
	20–24 years		25–35 years		36–45 years		Overall
Variable	Male $\bar{x}\,(\sigma)$	Female $\bar{x}\,(\sigma)$	Male $\bar{x}\,(\sigma)$	Female $\bar{x}\,(\sigma)$	Male $\bar{x}\,(\sigma)$	Female $\bar{x}\,(\sigma)$	$\bar{x}\,(\sigma)$
Time to 50 tweets (days)	166 (207.96)	249.75 (392.76)	271.31 (508.1)	107 (146.14)	103.11 (146.33)	169.4 (235.54)	175.72 (316.94)
Hashtag density (%)	34.10 (19.15)	30.22 (35.86)	36.96 (25.76)	51.12 (33.46)	23.15 (16.35)	21.66 (27.35)	37.17 (29.68)
Emoji density (%)	3.00 (2.67)	25.81 (29.38)	14.41 (14.08)	25.72 (26.15)	22.15 (22.84)	23.63 (24.54)	20.80 (22.51)
CTTR☆	15.06 (5.80)	9.37 (3.51)	13.39 (5.1)	12.22 (4.25)	14.79 (5.22)	10.48 (3.39)	12.56 (4.7)
Yule's K♝							

	54.54 (8.43)	107.1 (144.17)	50.87 (28.1)	51.82 (22.1)	52.52 (30.1)	59.03 (26.18)	58.73 (53.74)
Denn_density (%)	0.87 (1.25)	0.66 (0.95)	0.62 (1.00)	0.84 (1.30)	1.02 (0.87)	0.79 (0.95)	0.78 (1.05)
Weil_density (%)	0.64 (0.68)	2.44 (2.45)	0.89 (1.11)	1.73 (3.43)	1.21 (1.31)	0.73 (0.68)	1.36 (2.27)
Sentiment score$^{\infty}$	0.43 (0.20)	0.35 (0.26)	0.40 (0.25)	0.46 (0.22)	0.41 (0.25)	0.52 (0.18)	0.42 (0.24)

Note. Participants with no hashtags and/or no emojis were set to zero, and participants with fewer than 50 tweets (NAs) were removed for the computation of tt_50. The same was done for participants who did not have any ‘denn’ or ‘weil’ in their tweets. ☆Carrol’s corrected type token ratio; the higher the CTTR, the more lexical variety is in the text (Hess et al., 1984). ♞Yule’s K characteristic; the larger Yule’s K, the easier the text is to understand (Miranda-García & Calle-Martín, 2005). $^{\infty}$Sentiment scores are based on a -1, 0, +1 range (Novak et al., 2015b); *n* = 45, participants with sent scores. If a single tweet contained more than one emoji, the sentiment score for that tweet reflects the mean sentiment score of all emojis in this tweet.

Table 9 provides an overview of the variables to be used in hypothesis testing (excluding LIWC results). Time to 50 tweets is a measure that Shapp (2014) called ‘span,’ which is an indicator of how avid Twitter users are (the higher the number of days, the less avid). Interestingly, how avid a user is does not seem to depend solely on gender or age group but is rather an interaction between the two. Thus, while the 36–45-year-old male participants yielded the smallest number of days (103.11 on average) to reach 50 tweets, it was the 25–35-year-old female participants who produced the lowest number for the females (107 days on average) and were thus the most avid among the females. The females in the 20–24-year-age group took the longest with 249.75 days on average while the males in the 25–35-year age group took the longest with 271.31 days on average (the longest overall). When it comes to hashtag density (percentage of tweets with at least one hashtag in them), females in the 25–35-year-age group are clearly dominating the field with 50% of their tweets containing hashtags followed by the males in the same age group (~37%). Interestingly, in the 20–24-year-age group and the 36–45-year age group, males wrote slightly more tweets with hashtags on average than their female counterparts. The differences are, however, relatively small (20–24-year olds), or almost non-existent (36–45-year olds).

A similar picture presents itself for emoji density (percentage of tweets with at least one emoji in them) for the 25–35-year age group and the 36–45-year age group: while the females in the former seem to have quite a few more tweets with emojis on average (~26%), the difference is almost negligible in the latter suggesting that as age increases the difference in emoji density is diminished. The most striking difference occurred in the 20–24-year age group, in which ~26% of the female tweets had emojis in them on average. This is met with a diminishingly small percentage (3.0) for the male participants. Age does not play a discriminating role according to research (Hutchins, 2015, October 14), which seems to be the case here as well with the one very low outlier. Females seem to use more emojis on average, which is in alignment with current research as well (Bamman et al., 2014; Hutchins, 2015, October 14; SwiftKey, 2015, April 21b). When scrutinizing the numbers for the males, a reverse downward trend occurs with the oldest males having the most tweets with emojis and the youngest the fewest tweets with emojis on average. This sample in mind, this could cautiously hint at a current trend, in which younger males are moving away from using emojis in their tweets. The fact that ~20% of tweets contain at least one emoji confirms their important paralinguistic role to convey meaning that goes beyond the written word (Dresner & Herring, 2010; Kelly & Watts, 2015) and their ability to express feelings more accurately than words (Hutchins, 2015, October 14).

Overall, 2,844 emojis were captured by the emoji dictionary and an R-algorithm (Suárez-Colmenares, 2017). Male participants used 1,365 emojis (192 different emojis) in 1,054 tweets, while female participants used 1,489 emojis (208 different emojis) in only 895 tweets, nicely extending the insights on emoji density in Table 9 above. Women do not only seem to use more emojis, but they also use a more varied selection in fewer tweets. Table 10 below provides an overview of the top 15 emojis by gender, confirming the all-time-favorite and most-tweeted emoji, the face with tears of joy emoji, surpassing all other emojis for both males and females (Rothenberg, 2013a; Twitter Data, 2016, March 21).

Table 10: *Participants' Top 15 Emojis by Gender*

	Male			Female		
Rank	Description	Emoji	#	Description	Emoji	#
1	Face with tears of joy	😂	172	Face with tears of joy	😂	137
2	Grinning face with smiling eyes	😁	106	Smiling face with smiling eyes	😊	102
3	Smiling face with heart eyes	😍	94	Smiling face with open mouth & smiling eyes	😄	85
4	Smiling face with smiling eyes	😊	68	Smiling face with heart eyes	😍	83
5	Smiling face	☺️	67	Smiling face with open mouth	😆	54
6	Winking face	😉	63	See-no-evil-monkey	🙈	48
7	Green heart	💚	55	Red heart	❤️	45
8	Face with rolling eyes	🙄	48	Sun	☀️	42
9	Thinking face	🤔	33	Flexed biceps	💪	39

10	Face blowing a kiss	😘	28	Two hearts (Winking face)	💕 (😉)	34 (34)
11	See-no-evil-monkey	🙈	27	Smiling face (Victory hand)	☺️ (✌️)	31 (31)
12	Smiling face with open mouth & closed eyes (Smiling face with open mouth & smiling eyes)	😆 (😄)	23 (23)	Thumbs up	👍	27
13	Grinning face (Hot beverage)	😀 (☕)	22 (22)	Grinning face with smiling eyes	😁	24
14	Smiling face with sunglasses (Thumbs up)	😎 (👍)	16 (16)	OK hand	👌	23
15	Cloud with rain	🌧️	14	Face with rolling eyes (Flushed face)	🙄 (😳)	20 (20)

Let us now have a look at the two word-based measures in this study, the CTTR and Yule's K characteristic. It appears the males in this sample, on average, produced tweets, which have more lexical variety than the females' tweets across all age groups. Interestingly, the youngest males had the highest CTTR followed by the oldest male participants. For the females, it was the 25–35-year-olds who had the highest CTTR with the youngest females having the lowest — the situation is thus somewhat reversed. Conversely, female tweets, on average, seem to be easier to understand than male tweets across all age groups. Not surprisingly, the females in the youngest age group produced the highest Yule's K followed by the oldest females. One group seems to counter this trend: the youngest males had the highest CTTR, but also have the highest Yule's K among the males (see hypothesis testing below). This indicates that, while their tweets were lexically varied, the readability did not really suffer. Overall, this makes sense, as higher lexical variety can increase the difficulty of a text. A pattern emerges, which shows that the hashtag and emoji density might be connected to measures of lexical diversity.[40]

In terms of formality, the overall percentage of tweets containing *weil* 'because' was 1.36% compared to 0.78% of tweets containing *denn* 'because', which is the more formal of the two.[41] When looking at gender and age groups, more details about the distribution of *denn* and *weil* can be unearthed: While the males both in the 20–24-year and the 36–45-year age group seem to be 'more formal' than the women, the females in the 25–35-year age group are more formal. The females both in the 20–24-year and the 25–35-year age group seem to be more informal with higher percentages on *weil* than the males with the males only scoring higher in the 36–45-year age group. What is interesting is that the results for *weil* do not necessarily mirror this finding for *denn*: In the 25–35-year age group, the females scored higher both on *denn* and *weil*, while the males in the 36–45-year age group scored higher both for *denn* and *weil*. This seems to suggest that having high scores on both measures is not mutually exclusive and indicates that, for the age groups for which this is true, participants exhibit more lexical diversity.

Figure 16 below is in alignment with Figure 1, which shows that most emojis are on the positive end of the spectrum with seemingly miniscule differences in sentiment scores, as there are no sentiment scores lower than 0.3. Additionally, Figure 16 is a nice representation of

[40] Let us recall that the CTTR and Yule's K were measured on cleaned tweets without any emojis in them. However, when the user conceived of the tweet, they did so in context of hashtags and emojis, which is why this line of reasoning follows.
[41] *Da* 'because,' another formal variant, was excluded because it also serves as an adverb of place, which would have skewed the analysis.

participants' sentiment scores throughout the year aptly mirroring different levels of happiness dependent on the time of year. While sentiment scores are lowest in January (one of the coldest months in Germany), they steadily increase moving towards the more moderate months reaching a plateau in between roughly April and June, finally peaking in August (0.49). This development can most likely be attributed to the fact that, in addition to warmer temperatures starting around May, festivities such as Easter (April) and Pentecost (May) fall within those 'plateau' months. This is not so much important because of their religious meaning, but rather because they often coincide with off-days and vacation days for many people. June–August is high season for temperatures and vacations, which is why it is not surprising that sentiment scores are high during those months.

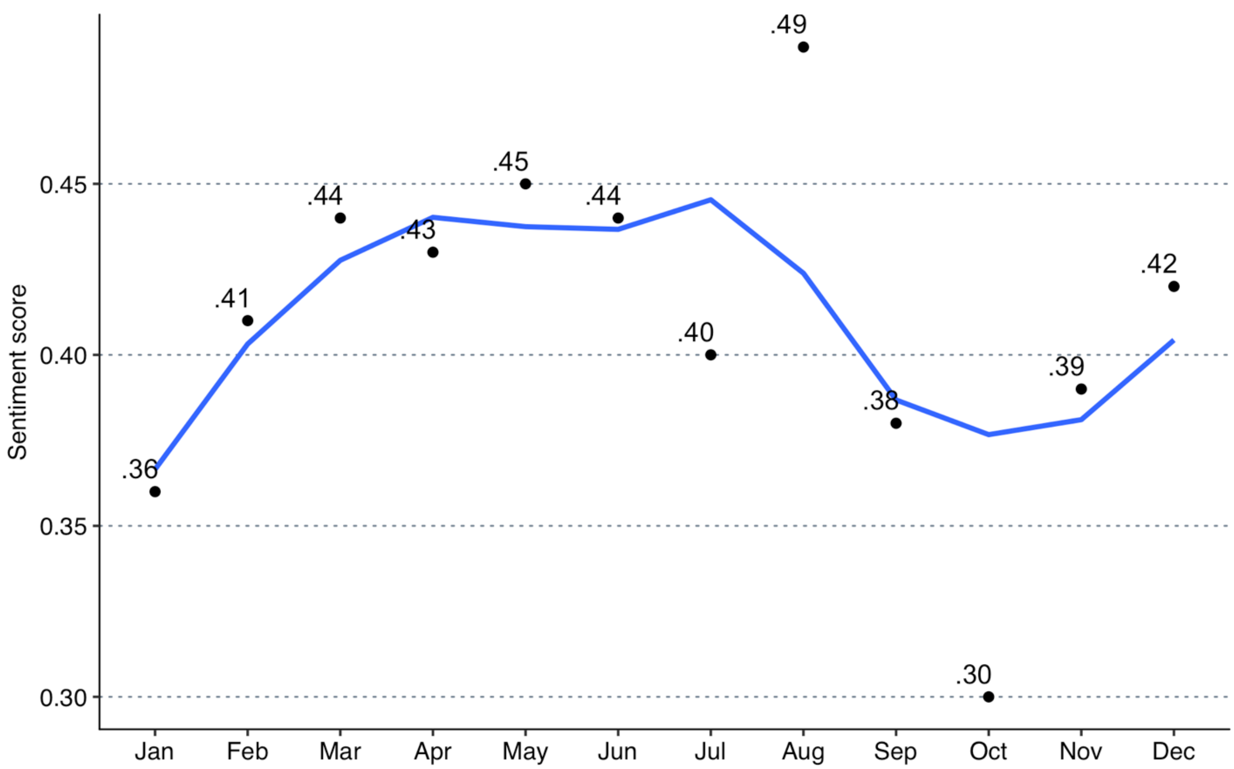

Figure 16. Participants' Mean Sentiment Scores throughout the Year.

The drop after July with its nadir in October probably stems from the fact that people have to return to work and temperatures are dropping with October and November being particularly gloomy in many parts of Germany. Sentiment scores pick up again towards the end of the year with holidays such as Christmas and New Year's Eve around the corner.

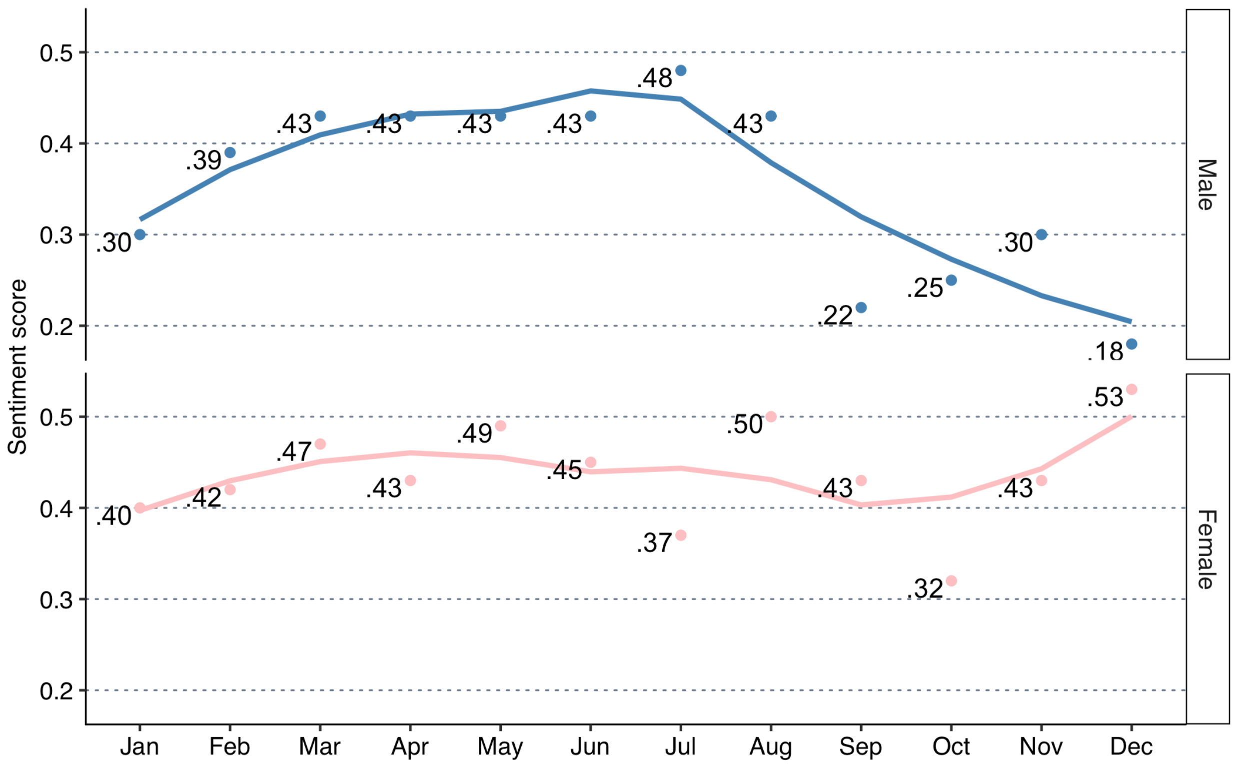

Figure 17. Participants' Mean Sentiment Scores throughout the Year by Gender.

Contrasting Figure 16 above with Figure 17, it becomes obvious that it is the males, who contribute to the sudden drop in sentiment scores in the months following July. In fact, the males maintain relatively stable scores throughout the months of March–June, peaking in July (.48), for reasons mentioned above, transitioning into a steady decline towards December with an unmatched low of .18. We can also infer that it is the women who are responsible for the sentiment scores picking up again towards the end of the year in Figure 17 with a score of .53, which is higher than the one in August (.50) while maintaining steady scores throughout the rest of the year.

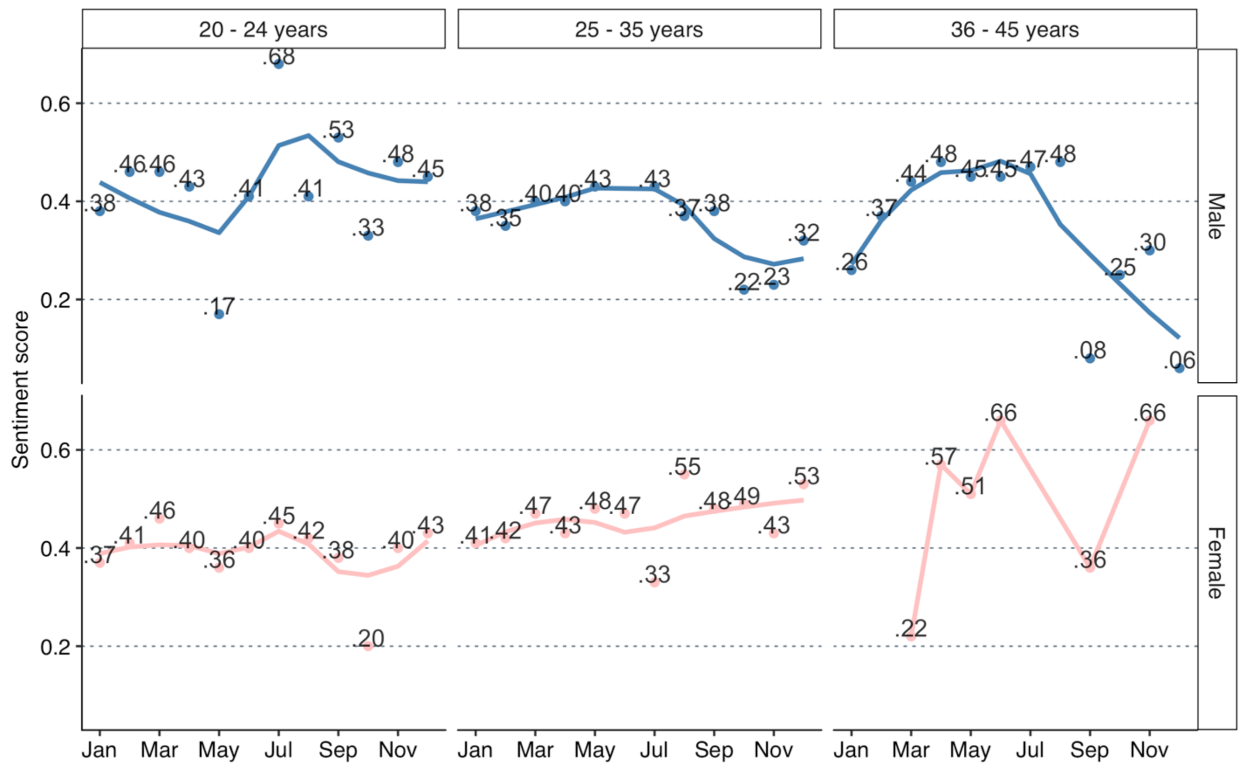

Figure 18. Participants' Mean Sentiment Scores throughout the Year by Gender and Age.

Once age is included in the visualization, the trends in Figures 16, and especially 17, are elucidated even more. We can infer that the females in the 36–45-year age group has fewer data points, with more variation between them. These outliers have the potential to distort the curve in Figure 16 to some extent. Due to the fact that both other age groups have data points throughout the year with relatively consistent, and less variable values, this trend is mitigated. As for the males, Figure 18 reveals that the dip toward the end of the year is mostly caused by the oldest male age group. Figures 16–18 exemplify the importance of visualizations beyond the surface level to reach a deeper understanding of trends in one's data, and what they are influenced by.

While hypotheses on punctuation are not explicitly part of this study, I would be remiss not to include my findings here before the actual hypotheses testing; for good measure, and because punctuation can be indicative of the level of formality of a given text (Thayer et al., 2010).

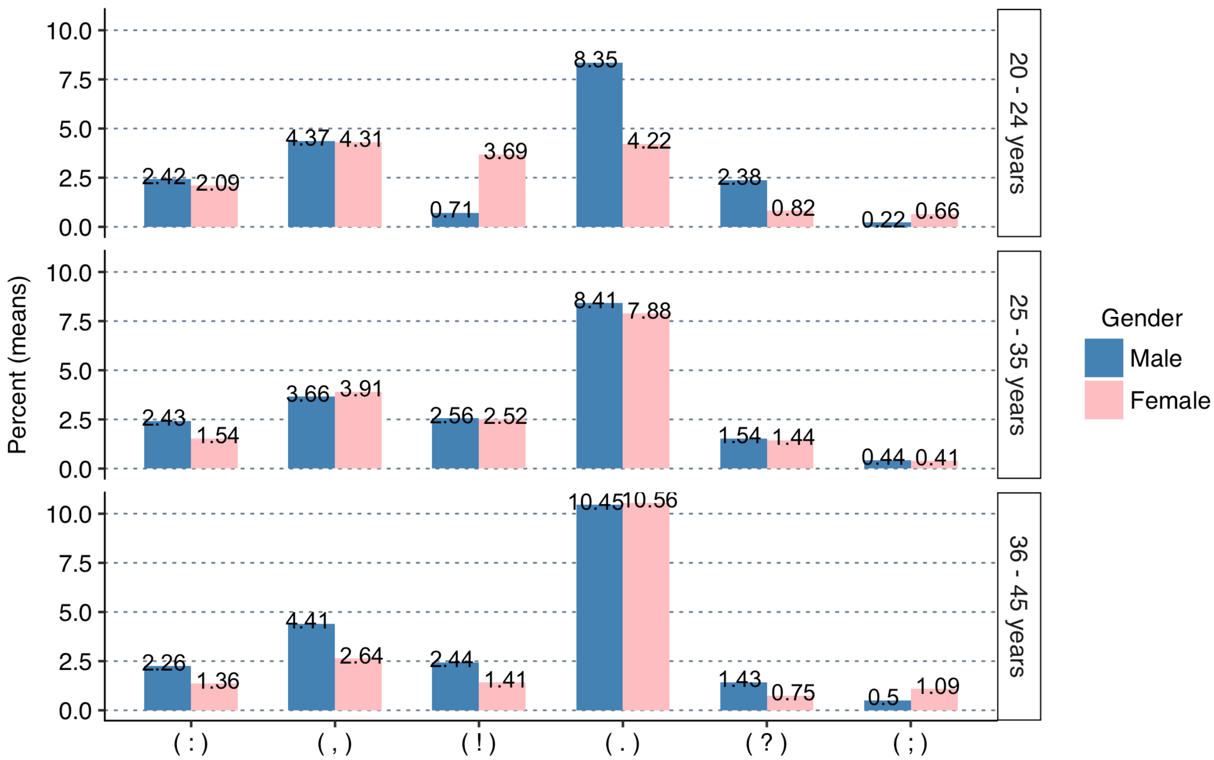

Figure 19. Participants' Use of Punctuation Marks [%] across all Tweets.

Inferring from Figure 19, it appears as if gender is not a discriminating factor when it comes to the percentage of tweets containing punctuation marks, and what these punctuation marks are. With exception of the period, all other punctuation marks are used rather sparingly; even the period does not exceed 11% on average — this finding warranted further investigation:

A multivariate multiple regression analysis was run with gender and age, as well as Big Five scores as covariates (IVs) and all punctuation marks (percentages) as outcome variables (DVs).[42]

Table 11: *Multivariate Multiple Regression Results: Punctuation*

	df	Approx *F*	num *df*	*p*
Age	1	2.45	6	0.037 *
Gender	1	1.75	6	
E	1	1.01	6	
A	1	4.43	6	0.001 **
C	1	1.01	6	
N	1	1.67	6	
O	1	1.01	6	

Signif. codes: 0.01 '**' 0.05 '*'

The results show that age as well as agreeableness (A) are significant predictors for the overall percentages of punctuation marks used by the participants. As Figure 19 already suggested, gender did not turn out to be a significant predictor.

Table 12: *Univariate Regression Follow-ups (Betas)*

	Period	Comma	Colon	Semicolon	Qu-mark	Exclam-point
Age*	.26**	.00 NS	-.00 NS	.02 NS	-.00 NS	-.08 NS
A**	2.12**	-.35 NS	-.72*	-.10 NS	.10 NS	1.23**

Signif. codes: 0.01 '**', 0.05 '*', not significant 'NS'

Since the multivariate regression showed that only age and agreeableness were significant, all other predictors were run in the model, but are not reported with the univariate follow-ups. The first model (period) was highly significant, $F(7, 54) = 3.69$, $p < 0.01$, $R^2_{adj} = .24$. The adjusted R-squared of .24 can be considered sufficiently large in a social science context

[42] **lm(formula = cbind(period, comma, colon, semic, qmark, exclam) ~ age + gender + e + a + c + n + o, data = punct)**.

(Abelson, 1985; King, 1986). The period-model shows that both age and agreeableness are significant predictors for the percentage of periods used by the participants. The colon-model was not significant overall, $F(7, 54) = 1.65$, $p = .14$, $R^2_{adj} = .07$, which is most likely attributable to the overall high number of predictors in the model compared to the overall $N = 62$.[43] The significance of agreeableness should therefore be taken with a grain of salt. The same is true for the model predicting the percentage of exclamation points, $F(7, 54) = 1.51$, $p = .19$, $R^2_{adj} = .05$. Again, agreeableness as a significant predictor of exclamation points should be taken with a grain of salt.

In sum, it appears that gender is not a significant predictor of the percentage of punctuation marks used in the tweets. This could indicate that males and females used punctuation marks to the same extent (percent). To reiterate, percentages were relatively low overall corroborating the finding that tweets are more oral-like and less formal. What is indeed interesting is that both age and agreeableness are significant predictors of the percentage of periods used (Table 12). As a punctuation mark that concludes a sentence, I venture that this is a good, and in this case, real measure of formality, as adding it requires a little more attention to detail and the conscious decision to 'complete' the sentence. The period is also less likely to be used repetitively like an exclamation point, for instance. Thus, this finding is neither surprising for age, indicating that the older the participant, the more they lean towards the standard, nor is it surprising for agreeableness, which, by its nature, is a measure of how agreeable people are, and thus, how they closely they might stick to explicit rules. Arguably, someone who uses a period is acting according to expected norms, grammar norms in this case, whereas a 'period-denier' is producing little statements of defiance with every tweet. This, however, might often happen

[43] Overall model significance 'considers' all effects together instead of individual coefficients.

below the level of consciousness. While periods were used the most in this study, Herring and Zelenkauskaite (2008) as well as Baron (2008) found that the omission of sentence final punctuation marks is the norm in text messages. This can certainly be extrapolated to tweets, which explains the overall low usage of sentence final punctuation marks, including question marks and exclamation points. Furthermore, we should not forget that punctuation in social media texts is often used to compensate missing paralinguistic cues and intonation present in face-to-face interaction (Werry, 1996). In addition, playfulness can result in the placement of multiple punctuation marks (e.g. !!!!), which is another sign for informality (Tseliga, 2007) or an expression of emotion (Luginbühl, 2003). It has also been argued that multiple punctuation marks are used "to enhance the readers' and writers' ability to experience the words as if they were spoken" (Danet, 2001, p. 17).

4.7 Hand-coded Hashtag Subset of Tweets

Filtering out all tweets with at least one hashtag produced a subset of 8,105 tweets, ~41% of all tweets, which exceeds Hong et al.'s (2011) findings (18% of German tweets contained hashtags in their sample). Of this subset, roughly 20% (1,621 tweets) were randomly sampled using R: 783 tweets from male participants and 838 tweets from female participants (diff = 55). This produced a hashtag data set comprising 2,666 hand-coded hashtags, males = 1,155; females = 1,511 (diff = 356). As mentioned above, this is bigger than Shapp's (2014) sample by a factor of ~1.63.

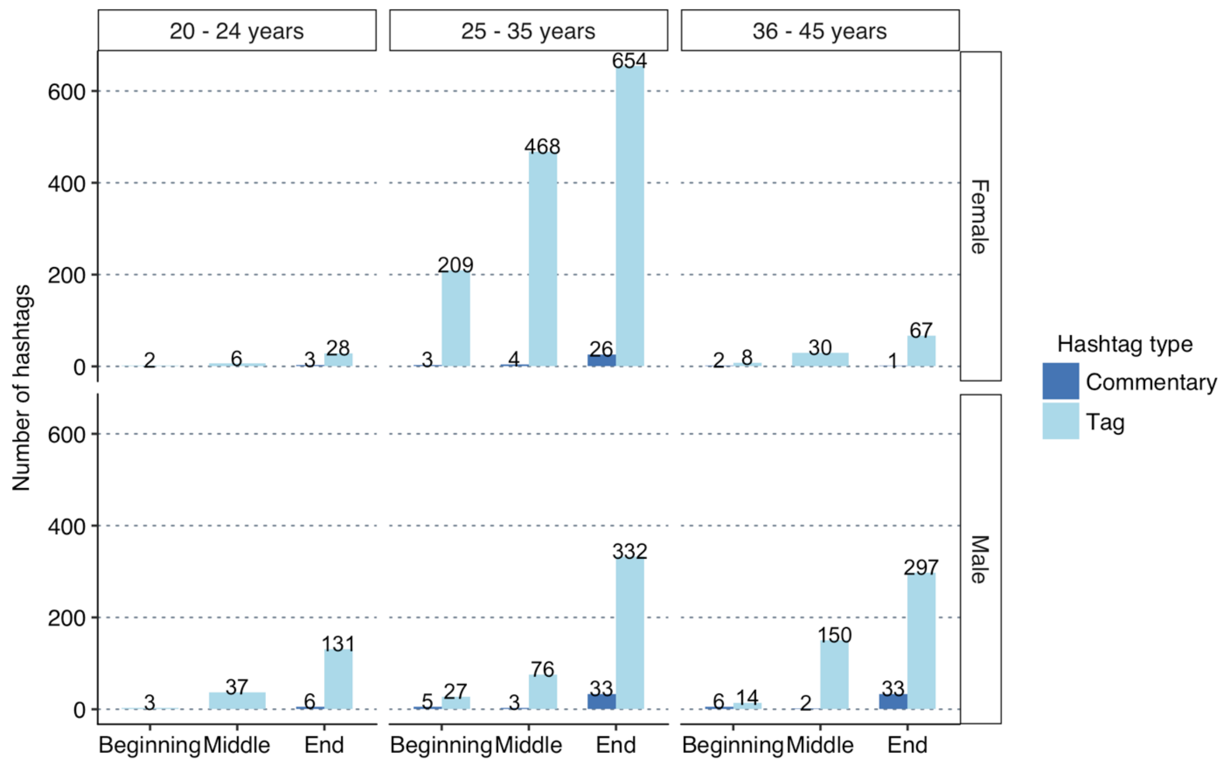

Figure 20. Participants' Hashtag Types and Position by Age and Gender.

To reiterate, tag hashtags are canonically used to tag topics, which is done by naming a concrete entity such as a person (#Obama, #Hillary), a place (#Paris, #TheHaven), a company (#Apple, #AMC), or an event (#Thanksgiving, #SummerOlympics). While these tags can be directed at different levels of the public ranging from tags relevant to tweets for the general public all over the world, or country, they also tag entities that are only relevant to certain individuals or a select group of Twitter users/followers (Shapp, 2014). Commentary hashtags, on the other hand, serve to "add additional meaning to the main semantic content of the tweet, and are not intended to practically connect the tweet to others that use the same hashtag" (Shapp, 2014, p. 7). Usually, commentary hashtags add an evaluation to what the author of the tweet just said. Routinely, this process adopts the following syntactic pattern: "Text body of the tweet in a sentence. #evaluation." For example, "When someone tells u its not safe to travel to a foreign place alone just cause ur female. #ridiculous #wearenotincapable."

Figure 20 elucidates that the tag-hashtag is used the most across gender, age group, and position in the tweet with the commentary hashtag used only very sparingly. This mirrors the answers given by the participants as to what they think they hashtag does in a tweet (Table 6). It is not too surprising that the end-position is favored across the board as it is the canonical position for hashtags, and tag-hashtags in particular (Twitter Inc., 2017g; Zappavigna, 2011). Interestingly, tag-hashtags are used liberally by females in the 25–35-year age group in all positions with the end position containing the most hashtags. Another interesting point to note is that male participants, while using mostly tag-hashtags as well (with the end-position also being favored), they maintain somewhat steady usage across age groups. This is in rather stark contrast to female participants, who display a lot of overall hashtag usage in the 25–35-year age group while the other two age groups use almost negligible amounts of hashtags. Overall, females produced 1,472 tag-hashtags versus males with 1,067 tag-hashtags. When it comes to commentary hashtags, the preferred position for all age groups is the end-position, which is not surprising given prior research (cf. Shapp, 2014). What is surprising, though, is that males produced more commentary hashtags (88) compared to females (39). This seems to indicate that the situation described by Herring and Paolillo (2006) and Shapp (2014) is reversed by virtue of the fact that men seem to be using less informational language (tag-hashtags) and more involved language (commentary), if only by a tiny margin. This issue will be revisited with hypothesis testing below to either corroborate or dispel this initial hunch. Overall, this almost perfectly mirrors participants answers to the question what they thought the function of a hashtag is. A clear majority (74.2%) of participants thought the hashtag is used exclusively for tagging purposes (see Table 6 above).

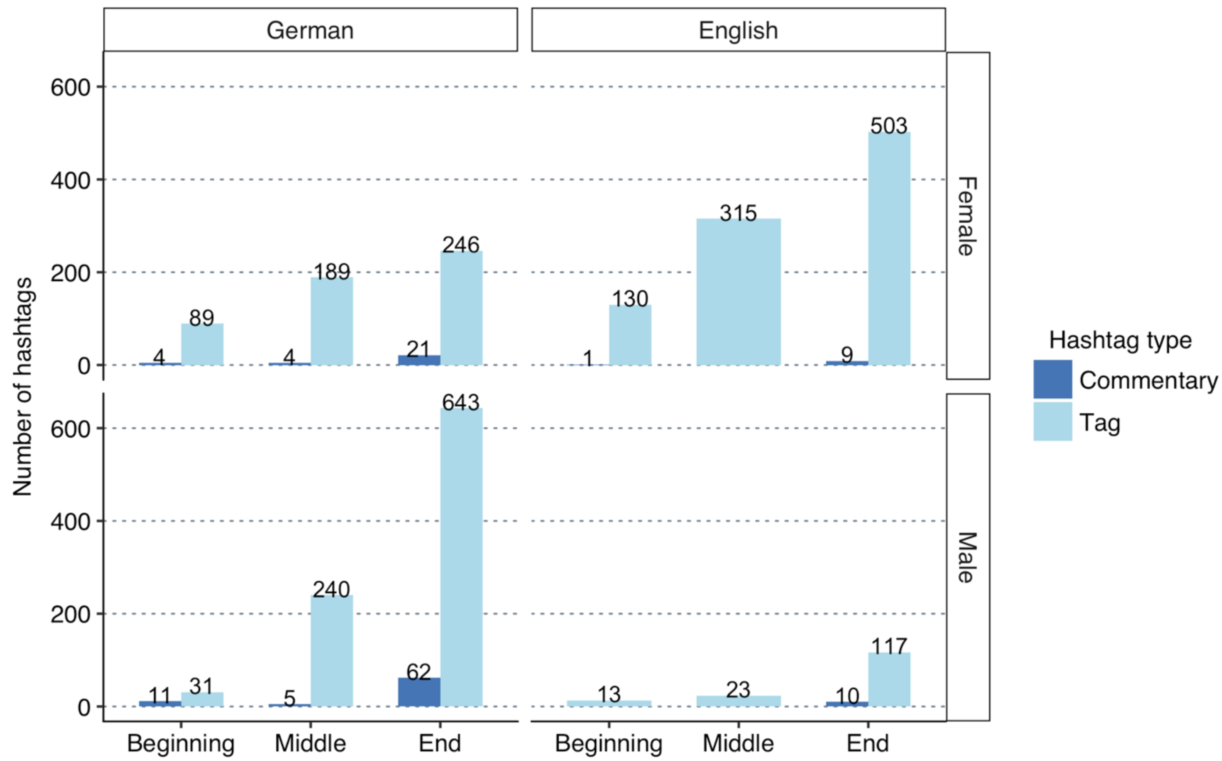

Figure 21. Participants' Hashtag Types and Position by Gender and Language.

Figure 21 yields two very interesting insights: (1) Among female participants, tag-hashtags are more equally distributed among both languages, 524 (de) vs. 948 (en) compared to males, 914 (de) vs. 153 (en). While males clearly favor German hashtags, the female participants in this sample are leaning towards a higher use of English hashtags. This could, among other things, suggest that the women are actively trying to be part of a larger audience, thus creating ambient affiliation (Zappavigna, 2011). Potentially, this might be linked to how they use Twitter as a professional tool, thus surpassing the geographic boundaries of German-speaking countries. Hashtags, especially tag-hashtags, have a classificatory function, which pertains to the topic of a tweet or its “aboutness” (Kehoe & Gee, 2011; Zappavigna, 2015), thus contributing to a more universal and/or global topic. The males on the other hand predominantly stick to German tag-hashtags thus creating ambient affiliation more limited to German-speaking countries and topics are arguable narrower in focus. In both languages, the end-position is clearly favored, mirroring the results from Figure 20 above.

(2) Comment-hashtags are predominant in German tweets with male participants using more (78) compared to female participants who used fewer (29). Both genders only produced ten English commentary hashtags. Again, in both languages, the end-position was the clear preference followed by beginning or middle position. As commentary hashtags are usually longer (sometimes having their own syntax) than tag-hashtags, their default position seems to be the end-position.[44] In addition, their overall lower frequency is attributable to the fact that they do not necessarily create ambient affiliation contributing to a larger topic, but rather commenting on, or replacing, the content of the tweet itself (Shapp, 2014; Zappavigna, 2011). Figure 22 below continues the story, but provides insight into the distribution between age groups. The interesting piece of information here is that the 25–35-year-old age group does not only contribute the most tag-hashtags (1,766), but also the most commentary hashtags (74) both in German and English. The second interesting finding is that it is also this age group that has the most English hashtags overall (999) compared to the 20–24-year-olds (27) and 36–45-year-olds (75). Bringing all three figures together, a picture presents itself that shows that the English tag-hashtag users are mostly 25–35-year-old females suggesting that they use them, at least in part, for professional purposes (given the age group). It is also conceivable that, as mentioned above with overall frequencies (they contributed over 10,000 tweets to the sample), the 25–35-year-

[44] In this study, commentary-hashtags that made up the entire tweet were coded as middle-position.

olds can be considered the early adopters of Twitter as a new technology over ten years ago. The 36–45-year-olds seem to be the laggards, and the 20–24-year-olds seem to have shifted mostly to other social media.

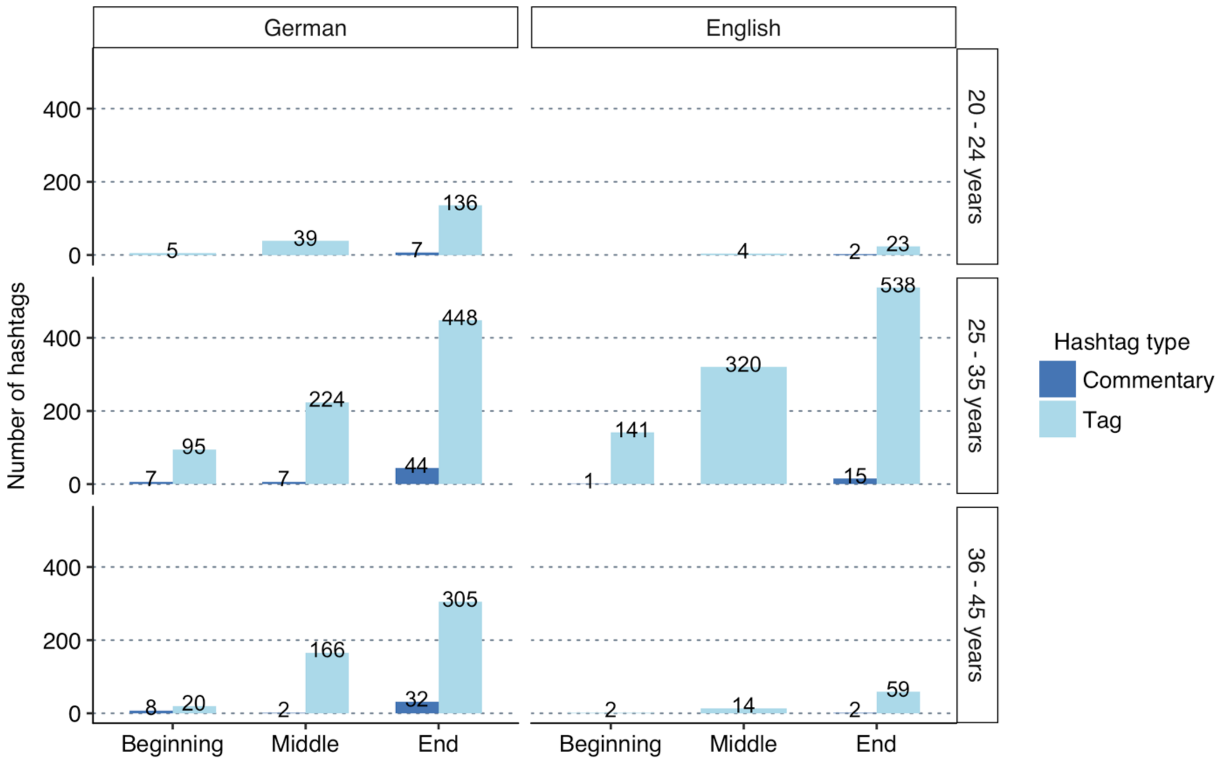

Figure 22. Participants' Hashtag Types and Position by Age and Language.

A note on hashtag position: While the end-position seems to be consistent as the canonical position both for tag- and comment-hashtags, the beginning and middle positions can (often) be considered part of the syntax and, as such, take on the experiential roles as defined by Halliday and Matthiessen (2004): processes, participants, and circumstances. Especially in middle or infix-position, tag-hashtags smoothly integrate into the syntax (Zappavigna, 2015). See examples (4a) and (4b) – shortened to protect privacy:

(4a) “#HOKO am 06.11.2014 in #München [...] ist dabei.“

#HOKO on 11/6/14 in #Munich [...] will be there

(4b) „[...] Kooperation mit #USA schwierig [...].“

[...] Cooperation with #USA difficult [...]

Examples (4a) and (4b) aptly illustrate how effortless all tag-hashtags are integrated in the tweet and become part of the syntax of the tweet (#HOKO seems to be an acronym used as a noun, and both #München and #USA function as nouns, or objects of their preceding prepositions). As Cunha et al. (2011) found, short hashtags have a better chance of being propagated. Thus, a quick look at the average length of both tag- and commentary-hashtags is in order: German hashtags have higher maximums (in characters), tag-hashtag: 44; comment-hashtag: 34, but also lower minimums, tag-hashtag: 1; comment-hashtag: 2. In comparison, English hashtags are shorter overall with lower maximums, tag-hashtag: 23; comment-hashtag: 2 vs. higher minimums, tag-hashtag: 2; comment-hashtag: 4. Table 13 adds some interesting insights to this finding: on average, German comment-hashtags are longer than tag-hashtags, but for English hashtags, there is almost no difference on average. In addition, the males have shorter German tag-hashtags than the females across all age groups while they, conversely, have longer comment-hashtags than the females. Looking at the English hashtags, the situation is less straight-forward: here, only the males in the 25–35-year age group have shorter tag-hashtags while only the males in the 36–45-year age group have shorter comment hashtags.

Table 13: *Average Hashtag Length (Char.) by Type, Gender, Age, and Language*

	Age groups						
	20–24 years		25–35 years		36–45 years		Overall
Hash-type	Male $\bar{x}\,(\sigma)$	Female $\bar{x}\,(\sigma)$	Male $\bar{x}\,(\sigma)$	Female $\bar{x}\,(\sigma)$	Male $\bar{x}\,(\sigma)$	Female $\bar{x}\,(\sigma)$	$\bar{x}\,(\sigma)$
German:							
Tag	6.80 (4.34)	7.19 (7.45)	7.93 (4.02)	10.36 (4.71)	6.09 (2.99)	9.28 (4.53)	8.02 (4.48)
Commentary	12.50 (2.38)	10.33 (3.06)	14.47 (6.55)	13.29 (7.62)	12.38 (3.62)	5.50 (4.95)	13.06 (5.78)
English:							
Tag	10.04 (2.27)	8.50 (3.70)	9.05 (3.53)	10.98 (4.17)	8.81 (3.51)	6.71 (1.87)	10.50 (4.18)
Commentary	12.50 (0.71)	n/a	10.29 (4.64)	9.88 (3.22)	10.00 (n/a)	11.00 (n/a)	10.35 (3.44)

4.8 Summary of Tweet Corpus and Statistics

Overall, 19,772 tweets were collected, 11,678 (59.1%) from males and 8,094 (40.9%) from females. Gender did not turn out to be a significant predictor, but age did. Most tweets were produced in the months between January and April/May, in between 2am and 8pm (hour of day is a significant predictor). On average, it took participants roughly 176 days to produce 50 tweets with roughly 37% of tweets containing at least one hashtag and roughly 21% of tweets containing at least one emoji. The overall Carroll's corrected type token ration (CTTR) was 12.56 and the overall Yule's K was 58.73. Participants overall used more informal *weil* 'because,' 1.36% of tweets compared to formal *denn* 'because,' 0.78%. The overall sentiment score turned out to be rather positive with 0.42 on a -1, 0, +1 scale. The top emoji for both genders turned out to be the face with tears of joy emoji. The hashtag-subset yielded 2,666 hashtags overall (males = 1,155 vs. females = 1,511), with the tag-hashtag being used most across age, gender, and position in the tweet. Interestingly, men used more commentary hashtags than females, which is quite apart from previous research (e.g. Shapp, 2014). In sum, these insights seem to suggest that participants' Twitter language is more informal, indicating that Twitter is more on the oral end of the oral-written spectrum — this is tested below for significance. Further, interesting trends emerge, especially regarding the use of tag and commentary hashtags, with the tag hashtag being the most prominent. Relating to the hypotheses and stipulations in this study, this indicates gender interactions with Twitter features that are quite different from what has previously been suggested (Shapp, 2014), especially when looking at the increased commentary hashtag use for males.

4.8 Hypothesis Testing

Before delving into hypothesis testing, it needs to be mentioned that, in this study, an alpha level of $\leq .05$ was considered significant for all statistical tests, indicating that there is maximally a 5% chance that the findings reported are due to mere chance for any given hypothesis.

4.8.1 Effects of Personality on LIWC Categories

(1a) There will be a significant positive correlation between an extraverted personality (as measured by the score for extraversion in the Big Five factor model) and the percentage of positive emotion words.

(1b) There will be a significant positive correlation between an agreeable personality (as measured by the score for openness in the Big Five factor model) and the percentage of positive emotion words.

(1c) There will be a significant negative correlation between an agreeable personality (as measured by the score for openness in the Big Five factor model) and the percentage of swear words.

(1d) There will be a significant positive correlation between a neurotic personality (as measured by the score for openness in the Big Five factor model) and the percentage of words in the anxiety category.

For the sake of completeness, Table 14 not only contains the word categories tested in the hypotheses (highlighted as such*), but the entire set of word categories included in LIWC. It provides an overview of all categories tested in any of the hypotheses pertaining to LIWC categories.

Table 14: *Summary of Correlations of Big Five Personality Scores and LIWC Categories*

Category	E	A	C	N	O
wc	--	-.15	.01	.01	-.21
wps	.12	-.17	.03	.09	.14
pronoun	-.03	.10	-.14	.11	.15
i	-.06	.07	-.19	.13	.22
we	.19	.14	.14	-.07	.18
self	-.01	.10	-.16	.11	**.26***
you	-.02	.05	-.03	.11	-.09
other	.09	.10	-.04	-.03	-.12
negate	-.03	-.09	-.19	-.04	.01
assent	**-.26***	-.19	.12	.18	-.10
article	.12	-.09	-.14	-.07	-.10
preps	-.06	.06	.07	-.09	--
numbers	.08	-.15	.10	-.12	-.23

affect	-.10	.16	.13	**.26***	.03
positive_emotion★	-.07	.21	.17	.20	.07
positive_feeling✤	**.28***	**.27***	**.38****	-.05	.06
optimism	.19	-.02	.09	.07	.02
negative_emotion★	-.10	-.12	-.13	.22	-.12
anxiety★	.02	.06	-.05	**.37****	-.24
anger	.03	-.01	-.09	.17	-.16
sad	.01	.01	.04	-.05	.10
cognitive_mechanism	-.11	-.13	-.21	.21	-.04
cause	.04	-.07	-.07	.02	.02
insight	.14	-.03	-.09	.12	-.06
discrepancy	-.11	-.24	-.20	.02	.11
inhibition†	.07	.02	.01	-.02	.11
tentative★	-.14	-.11	-.15	.03	-.01
certain	-.22	.01	-.14	.32	-.12
social★	-.34	-.04	-.15	.25	.18
communication	**-.08****	.07	--	**.03***	.08
other_reference	-.11	.13	.08	.12	-.02
friends★	-.23	-.03	.02	-.01	.21
family★	-.08	.12	-.03	.21	-.12
humans⧗	.05	-.07	-.01	.08	.03
time	.04	.07	.01	.05	-.03
past⧗	-.11	.06	.03	.22	**.26***
present⧗	.07	.11	-.01	-.05	-.10
future⧗	-.04	-.06	-.17	.21	-.07
space	-.02	**.31***	.08	.04	.03
up	**-.31***	.08	.01	.09	-.13
down	-.15	.09	-.16	.24	-.17
incl✂	.15	**.26***	.08	.08	-.04
excl✂	.08	.03	.05	-.14	.16
motion	.06	-.02	-.13	-.02	.14
occupation★	.08	.07	.12	-.03	.13
school	-.09	.11	.04	.10	.10
job★	**-.31***	-.17	-.24	.17	-.18
achievement★	-.03	-.13	.19	-.13	-.21
leisure	.12	-.08	.21	-.15	-.16
home	.17	-.07	.24	-.13	**-.25***
sports★	.14	-.04	.17	-.18	-.07
tv	--	-.07	.11	-.06	-.11
music	.10	-.03	-.03	-.06	-.10
money★	.13	-.18	.10	.03	-.10
metaphor	-.16	-.05	-.06	.07	-.13
religion	.21	.07	.17	-.15	-.01
death	**.29***	.15	-.17	**-.26***	.16

physical	-.14	.01	.06	-.02	-.21
body	.03	-.10	-.32	.14	.04
sex	.14	-.06	-.23	.09	.06
eat	-.16	-.05	**-.27***	.11	-.06
sleep	--	.03	-.03	.10	.23
grooming	-.01	.01	-.20	.13	.13
swear★	.18	.15	.17	.03	.17
non_fluency	-.18	-.12	-.22	.16	.15
fillers	.04	-.10	-.06	.10	**.31***
allpunc	-.14	.20	-.16	.13	.02
period	-.13	**.28***	-.08	.12	.02
comma	-.05	-.18	**-.32***	.17	.19
colon	.02	**-.34***	.04	-.15	-.12
semicolon	.08	-.01	-.03	-.16	.07
question mark	-.11	-.04	-.09	-.14	-.24
exclamation point	-.03	**.29***	.03	.13	.02

Notes. Signif. codes: 0.01 '**', 0.05 '*'; ★Category tested in hypotheses; ✤ Positive feeling is a subcategory of positive emotion still present in the German LIWC2001 dictionary, which has since been removed from the English LIWC dictionary (Pennebaker, Chung, et al., 2007); † Very low base rate category - removed from LIWC2015, but still present in German LIWC2001 dictionary; ⧖Updated to more inclusive categories in LIWC2015, but still present in German LIWC2001 dictionary; ✂Weak psychometrics and thus removed from LIWC2015, but still present in German LIWC2001 dictionary (Pennebaker, Booth, Boyd, & Francis, 2015).

Hypotheses (1a)–(1d) were tested together by running a Pearson product-moment correlation coefficient on the LIWC word categories (percentages) and the participants' scores on the individual Big Five domains. Table 15 below includes one correlation coefficient per Big Five factor for each of the four LIWC word categories for hypotheses (1a)–(1d).

Table 15: *Correlations of Big Five Personality Scores and LIWC Categories*

Category	E	A	C	N	O
positive emotion	-.07	.21	.17	.20	.07
positive_feeling	**.28***	**.27***	**.38****	-.05	.06
swear	.18	.15	.17	.03	.17
anxiety	.02	.06	-.05	**.37****	-.24

Table 15 reveals that hypotheses (1a) and (1b) cannot be confirmed for this data set because of a lack of significant correlations, $r = -.07$, $p = .57$, and $r = .21$, $p = .11$ respectively.

Positive emotion words have been shown to positively correlate significantly with extraversion and agreeableness in previous research (Golbeck, Robles, Edmondson, et al., 2011; Küfner et al., 2010; Mairesse et al., 2007; Mehl, 2006; Pennebaker & King, 1999; Yarkoni, 2010). While the correlations are not significant for this category, the sizes of the correlation coefficients follow the tendencies in previous studies using German (M. Wolf et al., 2008). According to Wolf (personal communication, September 13, 2017) lower correlations could "stay" non-significant depending on the sample and genre of the text. In addition, it is conceivable that, for the sample in this study, the positive feeling category is more homogenous despite its relative size being smaller than that of the positive emotion category. In contrast, a closely related, linguistically virtually the same, category, positive feelings, revealed significant positive correlations for extraversion ($r = .28$, $p = .03$), agreeableness ($r = .27$, $p = .03$), and conscientiousness ($r = .38$, $p \leq .01$). Wolf et al. (2008) list some example words for the positive emotions and the positive feelings category that can be considered to belong to exactly the same semantic category, e.g. *glücklich* 'happy.' Markus Wolf confirmed that there are no substantial differences between both categories, which is why they inherently correlate highly with each other; positive_feeling is a subcategory of positive_emotion, which means that almost all positive feeling words are contained in the positive emotion category (Wolf, M., personal communication, September 11, 2017). This ultimately led the developers to exclude the positive_feeling category altogether starting with the English LIWC2007 version (Pennebaker, Chung, et al., 2007, p. 8).

This is interesting for two reasons: (1) it shows how 'flexible' the word categories are, and (2) it also confirms what has been found in previous research albeit to a lesser extent. Mairesse et al. (2007) confirmed significant correlates of extraversion and words in the positive feeling category while Yarkoni (2010) confirmed significant correlates between extraversion and agreeableness and words in the positive feeling category. Further, it is interesting that conscientiousness was significantly positively correlated to positive feeling words. Since this has not been established in previous research, potential cross-cultural implications are conceivable, indicating that for German Twitter users, high scores on conscientiousness and a high percentage of positive feeling words are not mutually exclusive. Naturally, this could also be attributable to variability in the sample. Since positive feeling words are a subcategory of positive emotion words, it is not surprising that the former is the more dominant of two very similar word categories ($\bar{x}$ = 3.81%, σ = 2.00 vs. $\bar{x}$ = 0.36%, σ = .39). High correlations of Big5 categories with positive feeling words could indicate a higher usage of words in the subcategory. A generalized additive model (GAM) using the **`mgcv`**-package (Wood, 2006) was run with extraversion, agreeableness, and conscientiousness as predictors, controlling for age, gender, and emoji density, see model (1) below.[45]

(1)
```
gam(positive_feeling ~ E + A + C + s(Age, by = Gender, bs = "fs") +
s(Emoji_dens, by = Gender, bs = "fs"), data = liwc_use)
```

Both age and emoji density were entered as factor smooths as the non-linear pattern varies between genders and this variation has to be captured by the model. The function arguments "**`bs`**" controls the smoothing basis, and "**`fs`**" stands for factor smooth (also see Winter & Wieling, 2016). While interactions are usually not part of GAMs, it is possible to add an interaction term to the smooth with the **`by`**-argument (this can either be a categorical or a continuous interaction, which has to be mentioned explicitly in the model as a term since the

[45] The personality score variables (E, A, C, N, O) are ordinal in nature, but were treated as continuous in this and all following models since the spacing between values is exactly equal across all scores and domains (Pasta, 2009).

interaction terms have centering constraints applied to them (M. Clark, 2016, June 26)). The model was fit striking a balance between deviance explained (a generalization of R^2), the adjusted R^2, and the GCV[46] score to avoid overfitting. A Chi-square test for model comparison with a linear model revealed a significantly better fit of the GAM. Thus, the GAM was chosen over the LM with the same parameters because of its stronger explanatory power. Insignificant predictors were left in the model if they contributed to the explanatory power of the model without dropping them in a stepwise approach. Visual inspection of the residual- and QQ-plot did not indicate any problems in terms of homoscedasticity and normality.

Table 16: *Generalized Additive Model Results: Dependent Variable: Positive Feeling Words*

Fixed effects Parametric coefficients	*β*	*SE*	*t*	*p*
Intercept	-0.73	0.30	-2.45	$.02^{*}$
E	0.1	0.05	2.07	$.04^{*}$
A	0.06	0.06	1.07	$.29^{NS}$
C	0.16	0.06	2.77	$.008^{**}$
Approx. significance of smooth terms	*edf*	*Ref df*	*F*	*p*
s(Age):Gendermale	1.00	1.00	0.001	$.98^{NS}$
s(Age):Genderfemale	5.85	6.82	2.29	$.05^{*}$
s(Emoji_dens):Gendermale	1.00	1.00	3.25	$.08^{NS}$
s(Emoji_dens):Genderfemale	7.10	7.96	2.58	$.02^{*}$
$R^2_{adj} = .45$; Deviance explained = 61%				

Signif. codes: 0.01 '**', 0.05 '*', not significant 'NS'

Table 16 reveals significant fixed effects for extraversion (E) and conscientiousness (C), thus corroborating the finding above and showing that both extraversion and conscientiousness are significant predictors for positive feeling words while agreeableness (A) did not turn out to be a significant predictor. As in prior research, the relationship is of a positive kind (Mairesse et

[46] The generalized cross validation score (GCV) is similar to the AIC and can thus be used for model comparison. As with the AIC, lower scores are better (M. Clark, 2016, June 26).

al., 2007; Yarkoni, 2010): The percentage of positive feeling words increases by a factor of roughly 0.1 for every unit increase on the extraversion score, ceteris paribus (see also Figure 23 below), which implies that the more introverted the participants are, the fewer positive feelings words they use. For conscientiousness, for every one-unit increase in C, the percentage of positive feeling words increases by a factor of roughly 0.16%, ceteris paribus. Moving on to the significance of smooth terms: age turned out to be significant[47] (for females) as a non-linear predictor in interaction with gender. The effective degrees of freedom — to reiterate, an edf of ~1 indicates a linear pattern, and any edf > 1 indicates a non-linear pattern — reveal that the age and gender interaction is interesting as the males seem to follow a linear trend whereas the females follow a highly non-linear trend. In terms of emoji density, again, only the female interaction came back significant with a non-linear pattern while the male interaction is linear and non-significant. The visualization in Figure 23 below exemplifies a GAMs strength to capture non-linear relationships between predictors and outcome variables. Interestingly, the non-linearity in the interaction between gender and age and positive feeling words only applies to female participants in the sample (edf = 5.85) while there is a linear relationship for males (edf = 1.0). The first three variables (E, A, C) represent the linear portion of the GAM and the corresponding plots are residual plots like in a simple linear regression model. For the age variable, there are two distinct peaks of percentage of positive feeling words right around 30 years of age and then following a steep incline towards 45. First and foremost, this is similar to previous findings that found a non-linear relationship between age and affect also controlling for Big Five personality domain scores (Mroczek & Kolarz, 1998). It also seems to be in alignment with the puzzling relationship of age and happiness; still a somewhat contentious issue in

[47] P-values in GAMs are somewhat fuzzy, and, as Clark (2016) suggests, should not be used as definitive cut-offs at the .05 level, but can be interpreted to be significant if they are low, which is the case here.

psychological research. While there are several studies confirming a U-shape with varying age spans as for minimum level of life-satisfaction/happiness usually around 35–50 years (Blanchflower & Oswald, 2008; Jeste & Oswald, 2014; Stone, Schwartz, & Broderick, 2010), there are just as many studies that found linear increases or decreases or even flat trajectories throughout age (Charles, Reynolds, & Gatz, 2001; López, Møller, & Sousa-Poza, 2013; Mroczek & Kolarz, 1998). While the U-shape is narrower age-wise in this study, it seems to suggest that happiness/satisfaction (expressed via positive feeling words) increases toward older age following a non-linear pattern thus confirming one of the previous findings. I venture that these findings should, however, be taken with a grain of salt due to the contentious findings in prior research, and because the measure of positive feeling words is very likely not the only indicator for overall happiness/satisfaction. The interpretation of emoji_density in relationship to positive feeling words is straightforward in that there is a steady increase of positive feeling words in between 25–60% of tweets with emojis. This is not surprising and simply indicates that people who use a lot of emojis, also use a lot of positive feeling words. What is interesting though is the sharp drop with a minimum at 75% with a following increase. This could be attributable to the fact that people in that range of emoji density simply use fewer positive feeling words because they compensate for them with emojis, or it could also be an anomaly attributable to this data set, which would better explain the increase following 75% emoji density.

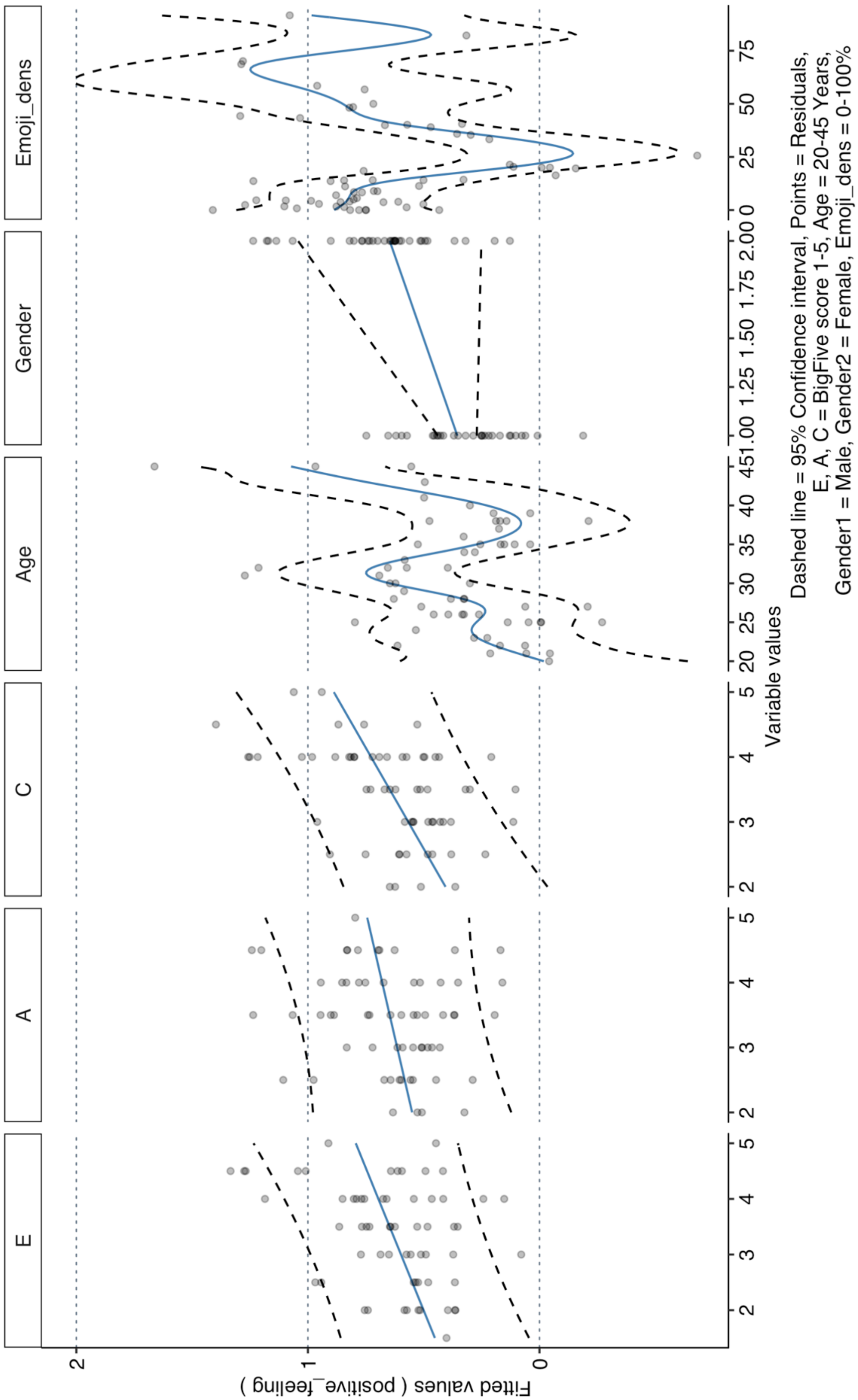

Figure 23. GAM-Plot: Positive Feeling ~ Predictors.

Hypothesis (1c) can be tested, again, by looking at Table 15 above. Since there is no significant correlation of any sort between agreeableness and swear words (r = .15, p = .24), hypothesis (1c) must be rejected thus opposing existing research, which found significant negative correlations between agreeableness and swear words (Holtgraves, 2011; Mehl, 2006; Yarkoni, 2010). This finding suggests a cross-cultural difference in that there seem to be general differences between how English and German are used, and how those differences relate to individual personality factors. In sum, the score on the Big5 agreeableness dimension does not seem to have any bearing on how many swear words the participants used. It is worth mentioning however, that the trend of the correlation between agreeableness and swear words is positive, and, while non-significant, this could indicate the use of swear words in the context of agreeableness as being used to build and maintain intimacy, union, and solidarity (Daly, Holmes, Newton, & Stubbe, 2004; Graziano & Eisenberg, 1997; McLeod, 2011), with the correlation remaining non-significant, which could be attributable to the size of the sample.

Hypothesis (1d) can be confirmed since there is a highly significant positive correlation between scores on the neuroticism dimension and the percentage of words related to anxiety, r = .37, $p \leq .001$. This result corroborates prior findings, which also found significant positive correlations for neuroticism and anxiety words (Gill et al., 2006; Nowson, 2006; Yarkoni, 2010).

The results pertaining to participants' scores on the Big Five dimensions and LIWC categories suggest that there is no homogenous alignment with previous findings. These individual differences could be due to general language dissimilarities between English and German. However, it is more likely that cross-cultural differences are in play, which indicate that German Twitter users' personalities map differently onto the usage of words in specific LIWC categories compared to English speakers (US). This shows in the significant positive correlations with positive feeling words, which present a subcategory of positive emotion words, thus, by proxy, confirming hypothesis (1a). Arguably, it would be too far-fetched to simply establish a cause-effect relationship between German language users, their personalities, and the frequencies, with which they use words from LIWC categories compared to English (mostly US) users of Twitter and their personalities. However, since the Big Five inventory's consistency has not only been established across cultural lines (John, 1990; R. McCrae & Costa, 1990), but also across different languages (Digman, 1990; R. McCrae & John, 1992), it is conceivable that there exist real cultural differences tied to how different personality types use a language that is not English in a different cultural environment.

> (1e) There will be a significant prediction of the sentiment score (sentiment scores for emojis range from -1 to +1, with 0 being neutral (Novak et al., 2015b)) by extraversion (as measured by the score for extraversion in the Big Five factor model).
> (1f) There will be a significant prediction of the sentiment score (sentiment scores for emojis range from -1 to +1, with 0 being neutral (Novak et al., 2015b)) by neuroticism (as measured by the score for neuroticism in the Big Five factor model).

To test hypotheses (1e) and (1f), a tweet subset was created containing only tweets, for which a sentiment score could be calculated due to the presence of emojis n = 1,789. This also resulted in the reduction of participants for this part of the analysis n = 45. As a second step, a linear mixed effects regression analysis was performed using the **`lme4`**-package (Bates, Maechler, & Bolker, 2015) in R to resolve the non-independencies[48] in the subset. This was also

[48] The assumption of independence of observations is one of the most, if not the most, important assumption for most statistical analyses.

tested using a simple LM, which resulted in all significant results elucidating what happens when the assumption of independence is violated and said violation is not accounted for appropriately in the model. Since individual users produced more than just one tweet and every one of their tweets received a specific sentiment score, these random effects have to be captured by the statistical model. To do that, I entered extraversion, neuroticism, age, gender, and emoji_density as fixed effects. As random effects, intercepts for participants were included, as well as by-participant random slopes for the effects of age and extraversion, see model (2) below:

(2)
```
lmer(sent_score ~ e + n + age + gender + emoji_dens + (1 + e +
age|part_id), data = sent_data, REML = FALSE)
```

and, mutatis mutandis for neuroticism, model 3:

(3)
```
lmer(sent_score ~ n + e + age + gender + emoji_dens + (1 + n +
age|part_id), data = sent_data, REML = FALSE)
```

Visual inspection of the respective residual- and QQ-plots did not indicate any problems in terms of homoscedasticity and normality. Since p-values are not as straightforward in mixed models, and, in this context, are not the be-all and end-all (Johannson, 2011), p-values for model fit were obtained running likelihood ratio tests of the respective full models with the effect (E), against the null model leaving out the effect in question (E), and mutatis mutandis for N. Model (2) revealed that extraversion affected the sentiment score significantly, $\chi^2(1) = 4.07$, $p = .04$, increasing it by .04 ± .02 (*SE*), on average ceteris paribus. Hypothesis (1e) can therefore be confirmed. Several lines of research have converged on the conclusion that high scores on extraversion have a significant relationship with happiness/positive sentiment or subjective well-being (Oerlemans & Bakker, 2014; Pavot, Diener, & Fujita, 1990; Salary & Shaieri, 2013). The finding at hand thus corroborates previous research and sheds light on the sentiments produced by German extraverts (males and females) on Twitter via their emoji usage. Further, in the case of German Twitter users, the rich-get-richer hypothesis, i.e. extroverts transfer their offline sociability to social media, rather than the social compensation hypothesis, i.e. introverts gain from social media, seems to be supported here (Correa et al., 2010; Valkenburg & Peter, 2007). As Oerlemans and Bakker (2014) claim, extraverts' happiness can be attributed to the fact that they participate more in social activities, which can be rewarding in themselves. I venture that this can be extrapolated to social media activities, which serve as a surrogate for actual face to face social activities, during which the reciprocal placement and the reception of emojis is rewarding.

Model (3) furnishes evidence that neuroticism did not affect the sentiment score significantly, $\chi^2(1) = 1.16$, $p = .29$; hypothesis (1f) must thus be rejected. While previous research has shown a significant negative relationship between happiness and neuroticism (Salary & Shaieri, 2013), this cannot be confirmed here based on the sentiment scores that were produced by the participants in this study through their emoji usage. For good measure, I also tested whether gender and age were significant predictors of sentiment scores. To do that, I entered purely demographic variables, age, gender, relationship status, and education2 (university education vs. no degree), as fixed effects. As random effects, intercepts for participants were included, see model (4):

(4)
```
lmer(sent_score ~ age + gender + relationship + edu2 + (1|part_id),
data = sent_data, REML = FALSE)
```

Model (4) did indeed reveal that gender affected the sentiment score significantly, $\chi^2(1) = 3.85$, $p = .05$, increasing it by .07 ± .03 (*SE*) on average, when switching from male to female ceteris paribus (confirming what Figure 17 above implied). This finding can be attributed to the fact that (1) women in this study used more emojis overall, (2) that they have been found to be happier overall and across cultures (Zweig, 2015), and (3) more frequent use of positive emojis, especially smiley emojis, has been linked to higher scores on agreeableness, conscientiousness, and openness (Wall et al., 2016), which is reflected in the female participants' higher scores on agreeableness and openness (cf. Figure 10 and Table 7). However, the same model testing for the significance of age revealed that age did not affect the sentiment score significantly, $\chi^2(1) = 0.21$, $p = .65$.

Since the mean sentiment scores are all positive, it is conceivable that the Pollyhanna Hypothesis (Boucher & Osgood, 1969) has to be credited here, at least in part. According to the the aforementioned hypothesis, "there is a universal human tendency to use evaluatively positive words […] more frequently and diversely than evaluatively negative words" (Boucher & Osgood, 1969, p. 1) during communication. This means humans usually focus on the positive things in life. I venture that this can be applied to a social media context as well, in particular to positive/negative emojis, which explains the overall positive sentiment scores in this sample (cf. Table 10; top 15 emojis for males and females are all positive).

4.8.2 Summary/Discussion of First Set of Hypotheses

While significant correlations between extraversion and positive emotion words and agreeableness and positive emotion words could not be confirmed, I found significant positive correlations for extraversion, agreeableness, and conscientiousness and the semantically very closely related category, positive feeling. Conceivably, this disparity is at least in part attributable to how the German LIWC2001 dictionary was conceived (M. Wolf et al., 2008). This means that the rejection of the above hypotheses should not be taken at face value, but rather with the newly found insights in mind. There was also no significant negative correlation between an agreeable personality and the percentage of swear words. However, as mentioned above, the trend of the correlation, while non-significant, is positive, potentially indicating a positive relationship between agreeableness and swear words, which are often used to build solidarity (Daly et al., 2004; Graziano & Eisenberg, 1997; McLeod, 2011). Corroborating previous findings, hypothesis (1d) was confirmed since a significant positive correlation was found between neuroticism and percentage of words related to anxiety. In terms of sentiment scores, hypothesis (1e) was confirmed indicating that higher scores on the extraversion dimension did indeed significantly increase sentiment scores. Neuroticism did, however, not significantly predict sentiment scores. An additional test for gender and sentiment scores revealed that gender is a significant predictor for sentiment score increasing it when switching from male to female.

Overall, these findings are partly in agreement with previous research, and partly quite apart from it. This is not really surprising in light of the disparities in previous research pertaining to personality traits and LIWC measures (for overview, see table in Qui et al., 2012, pp. 712-713).

4.8.3 Gender Effects and Twitter Measures

(2a) There will be a significant prediction of hashtag density (percentage of tweets containing hashtags) by gender.

Hypothesis (2a), which pertains to the overall prevalence of tweets with at least one hashtag, was tested running a linear model with age, gender, relationship status, and education2 as fixed effects. The LM was chosen over the GAM, because none of the smooth terms were significant, see model 5:

(5) `lm(hash_dens ~ age + gender + relationship + edu2, data = diss_data)`

Table 17: *Linear Model: Dependent Variable: Hashtag Density*

Fixed effects	β	*SE*	*t*	*p*
Intercept	124.43	40.12	3.10	.003**
Age	-1.77	0.65	-2.73	.009**
Genderfemale	3.72	7.63	0.49	.63NS
relationshipIn a relationship (2)	-44.32	30.63	-1.44	.15NS
relationshipSingle (3)	-38.27	30.31	-1.26	.21NS
relationshipMarried (4)	-9.84	31.29	-0.31	.75NS
edu2University of applied (2) sciences	13.32	13.67	0.97	.33NS
edu2University (3)	8.79	10.59	0.83	.41NS
edu2No degree (4)	-12.27	12.68	-0.97	.34NS
$R^2 = .23$; $R^2_{adj} = .11$				

Signif. codes: 0.01 '**', 0.05 '*', not significant 'NS'

The model results, summarized in Table 17, indicate that the overall model was not significant, $F(8, 53) = 1.97$, $p = .07$, which is why the significant p-values of the intercept and age cannot be trusted. The models overall lack significance, and thus explanatory power, and the absence of statistically significant results for gender resulted in hypothesis (2a) being rejected. These findings are probably attributable to the small sample size. Age does show a general tendency of older people using fewer tweets that contain hashtags, which is in alignment with the overall findings in this study. See the following hypothesis tests for more detailed analysis on the hand-coded hashtags subset of tweets.

(2b) There will be a significant prediction of hashtag type (tag vs. commentary – as measured by individual hashtag densities) by gender.

When testing hypothesis (2b), a similar situation as with the sentiment score above presented itself, as the hand-coded hashtag data set contained more than just one observation per participant. To capture these non-independencies, a generalized linear mixed effects model was run using the **lme4**-R-package (Bates et al., 2015). In addition to hashtag type (commentary vs. tag) as the dependent variable, I entered gender, age, language, position, and length of hashtag (nchar.) as fixed effects. As random effects, intercepts for participants were included, see model 6.

(6)
```
glmer(hash_type ~ gender + age + lang + pos + nchar + (1|part_id),
                  data = hash_data, family = binomial(link = logit))
```

P-values for model fit were again obtained running a likelihood ratio test of the full model with the effect (gender), against the null model leaving out the effect in question, $\chi^2(1) = 0.74$, $p = .39$. The results of model 6 thus indicate that gender did not affect the choice of hashtag type significantly; hypothesis (2b) was thus rejected as well. Contradicting current research (Shapp, 2014), which found that males use more tag-hashtags (more "informational") compared to females, who use more commentary-hashtags (more "involved"). The finding at hand indicates that gender does not play a significant role when it comes to the selection of the hashtag type. German Twitter users, both males and females, seem to use both types indiscriminately for knowledge management (tags) and self-expression (commentary) (Herring & Paolillo, 2006; Shapp, 2014), thus counteracting the notion of gender enactment through different types of hashtags. It is, however, conceivable that the topic of the tweet or the hashtag conveys gender, but this warrants further investigation. For example, Cunha et al. (2012) do not distinguish tag and commentary hashtags, but rather personal involvement-hashtags, written in 1st person singular versus persuasion-hashtags using the 3rd person imperative form (in Brazilian tweets). They claim that males use more persuasive hashtags compared to females, but also had to concede that most hashtags are neutral with only some indicating gender at all.

This does not only show that existing research is contentious, but also that cross-cultural differences are likely in play here, which could be responsible for the different usage behaviors of participants in this sample and Shapp's (2014) and Cunha et al.'s (2012) findings.

> (2c) There will be a significant prediction of hashtag type (tag vs. commentary – as measured by individual hashtag densities) by language (de vs. en).

Hypothesis (2c) was tested with the same data set and with the same model (6) as hypothesis (2b), mutatis mutandis, for the null model. The likelihood ratio test of the full model with the effect (language) against the null model leaving out the effect in question (language) came back non-significant, $\chi^2(1) = 1.88$, $p = .17$. This shows that language did not affect the choice of hashtag type significantly; hypothesis (2c) had to be rejected as well. Interestingly, German Twitter users seem both languages indiscriminately for both hashtag types refuting the assumption made in Chapter 2 that an English tag-hashtag in a German tweet would be more likely be used for ambient affiliation (Zappavigna, 2011); in fact, both languages were used for tag-hashtags to contribute to a searchable topic and a wider online audience that surpasses the borders of German-speaking countries.

Table 18: *Generalized Linear Mixed Effects Model: Dependent Variable: Hashtag Type*

Fixed Effects				
	β	*SE*	*z*	*p*
Intercept	3.51	1.69	2.08	.04*
genderMale	-0.54	0.63	-0.86	.39^{NS}
age	0.06	0.05	1.10	.27^{NS}
langen	0.43	0.32	1.35	.18^{NS}
posMiddle	1.76	0.52	3.40	.0006***
posEnd	-0.06	0.38	-0.16	.87^{NS}
nchar	-0.21	0.02	-8.92	2e-16***
Odds ratios		95% CI		
	β	lower	upper	
genderMale	0.58	0.17	2.01	
age	1.06	0.96	1.17	
langen	1.53	0.83	2.84	
posMiddle	5.79	2.11	15.98	
posEnd	0.94	0.45	1.97	
nchar	0.81	0.77	0.85	
$R_M^2 = .22;\ R_C^2 = .59$				

Note. Signif. codes: 0.001 '***', 0.05 '*', not significant 'NS.' Marginal R, R_M^2, pertains to the variance explained by the fixed effects while conditional R, R_C^2, can be interpreted as the variance explained by the whole model (fixed and random effects) (Nakagawa & Schielzeth, 2013).

Looking at the results of model 6 in Table 18, we learn that both the middle position (part of the syntax) and the length of the hashtag are significant predictors for the hashtag type. Compared to a hashtag in the beginning position, a hashtag in middle position increases the propensity of the outcome to be of type tag by 1.76 units on an unknown scale, while a one unit increase in the length of the hashtag (one character) decreases the propensity for the outcome to be of type tag by 0.21 units on an unknown scale, ceteris paribus. The odds ratio is usually more straightforward than the log-of-odds: compared to the beginning position, a hashtag in middle position increases the odds of a tag-hashtag by a factor of 5.79 (479%)[49] while a one unit

[49] Percentage = |Odds ratio - 1|*100.

increase in length (character) decreases the odds of a tag-hashtag by a factor of 0.81 (19%). Both of those results make sense: due to their (shorter) length, tag-hashtags can be integrated into the syntax of a tweet more easily, while the longer the hashtag becomes, the more likely it is that is a commentary hashtag thus confirming the descriptive overview presented in Table 12 above.

For expository purposes, a negative binomial model with by-participant intercepts as random effects was included below (for the hand-coded subset), model 7. A likelihood ratio test comparing model (7) to a zero-inflated model with the same predictors showed that there was no statistically significant difference, $\chi^2(1) = 0, p = 1$. Model 7 was chosen because its AIC was slightly lower than that of the zero-inflated model.

(7)
```
glmmadmb(n ~ age + gender + hash_type + lang + pos + (1|part_id), data
= hash_data_n, family = "nbinom", zeroInflation = FALSE)
```

Table 19: *Negative Binomial Model: Dependent Variable: Number of Hashtags*

Fixed Effects	β	*SE*	*z*	*p*
intercept	-1.58	0.90	-1.75	$.08^{NS}$
age	0.03	0.03	1.27	$.21^{NS}$
gendermale	0.07	0.34	0.20	$.84^{NS}$
hash_typet	1.54	0.21	7.48	$7.3e\text{-}14^{***}$
langen	-0.49	0.17	-2.86	$.004^{**}$
pose	1.33	0.22	6.10	$1.1e\text{-}09^{***}$
posm	0.70	0.23	3.00	$.003^{**}$

Signif. codes: 0.001 ‘***’, 0.01 ‘**’, not significant ‘NS.’

Some interesting insights are to be gleaned from Table 19: (1) both gender and age are non-significant, confirming the findings from model 6 above. (2) The hashtag type is indeed a significant predictor of the number of hashtags produced; compared to commentary-tag, the expected log-count for tag-hashtags is higher by a factor of 1.54 on an unknown scale, ceteris paribus. (3) While language did not turn out to be a good predictor of the hashtag type, it is, in fact, a significant predictor for the number of hashtags produced; compared to German hashtags, the expected log-count for English hashtags is lower by a factor of 0.49 on an unknown scale, ceteris paribus. This indicates that the language is a good predictor of frequency, but not of hashtag-type confirming the findings from above (Table 17 vs. Table 18). (4) Finally, compared to the beginning position, the expected log-count for the end-position is higher by a factor of 1.33 on an unknown scale, and the expected log-count for the middle-position is higher by a factor of 0.70 on an unknown scale, ceteris paribus. This confirms the tweet-end position as the canonical position for hashtags of both types while, at the same time, confirming previous findings that hashtags are used in the beginning of tweets (e.g. to introduce a topic/theme) (Kehoe & Gee, 2011), or in the middle, as part of the syntax, where they fulfil the role of sentence constituents and meta-data (e.g. geographic location) at the superordinate level (Zappavigna, 2015). It further furnishes evidence for a more content-related tweet behavior for both male and female German Twitter users, underscoring the importance of the tweet-language, and the impact it has on how Twitter is used (Hong et al., 2011).

(2d) There will be a significant prediction of emoji density by gender.

Hypothesis (2d) was tested using a generalized additive model with gender, age, relationship status and education2 as parametric predictors, and time on twitter (min), emoji density as smooth terms as non-parametric predictors with overall hashtag density as the dependent variable, see model 7:

(7)
```
gam(emoji_dens ~ e + a + c + s(age, by = gender) + gender +
relationship + edu2 + time_twitter + s(hash_dens, by = gender), data =
diss_data)
```

Table 20: *Generalized Additive Model Results: Dependent Variable: Emoji Density*

Fixed effects Parametric coefficients	*β*	*SE*	*t*	*p*
Intercept	-14.24	33.51	-0.46	0.67
E	10.29	3.52	2.92	.006**
A	0.91	4.16	0.22	.83^{NS}
C	-0.20	4.72	-0.04	.97^{NS}
relationshipIn a relationship (2)	-9.15	25.97	-0.35	.73^{NS}
relationshipSingle (3)	4.79	25.13	0.19	.85^{NS}
relationshipMarried (4)	-15.24	25.73	-0.60	.56^{NS}
edu2University of applied (2) sciences	0.02	10.50	0.00	.99^{NS}
edu2University (3)	4.79	8.55	0.56	.59^{NS}
edu2No degree (4)	-17.17	10.28	-1.67	.10^{NS}
genderfemale	10.20	6.58	1.55	.13^{NS}
time_twitter_min	0.37	0.33	1.14	.26^{NS}
Approx. significance of smooth terms	*edf*	*Ref df*	*F*	*p*
s(age):gendermale	2.59	3.24	3.25	.03*
s(age):genderfemale	2.01	2.47	3.33	.04*
s(hash_dens):gendermale	1.00	1.00	0.04	.84^{NS}
s(hash_dens):genderfemale	7.83	8.35	2.20	.05*
$R^2_{adj} =$ .35; Deviance explained = 60.9%				

Signif. codes: 0.01 ‘**’, 0.05 ‘*’, not significant ‘NS’

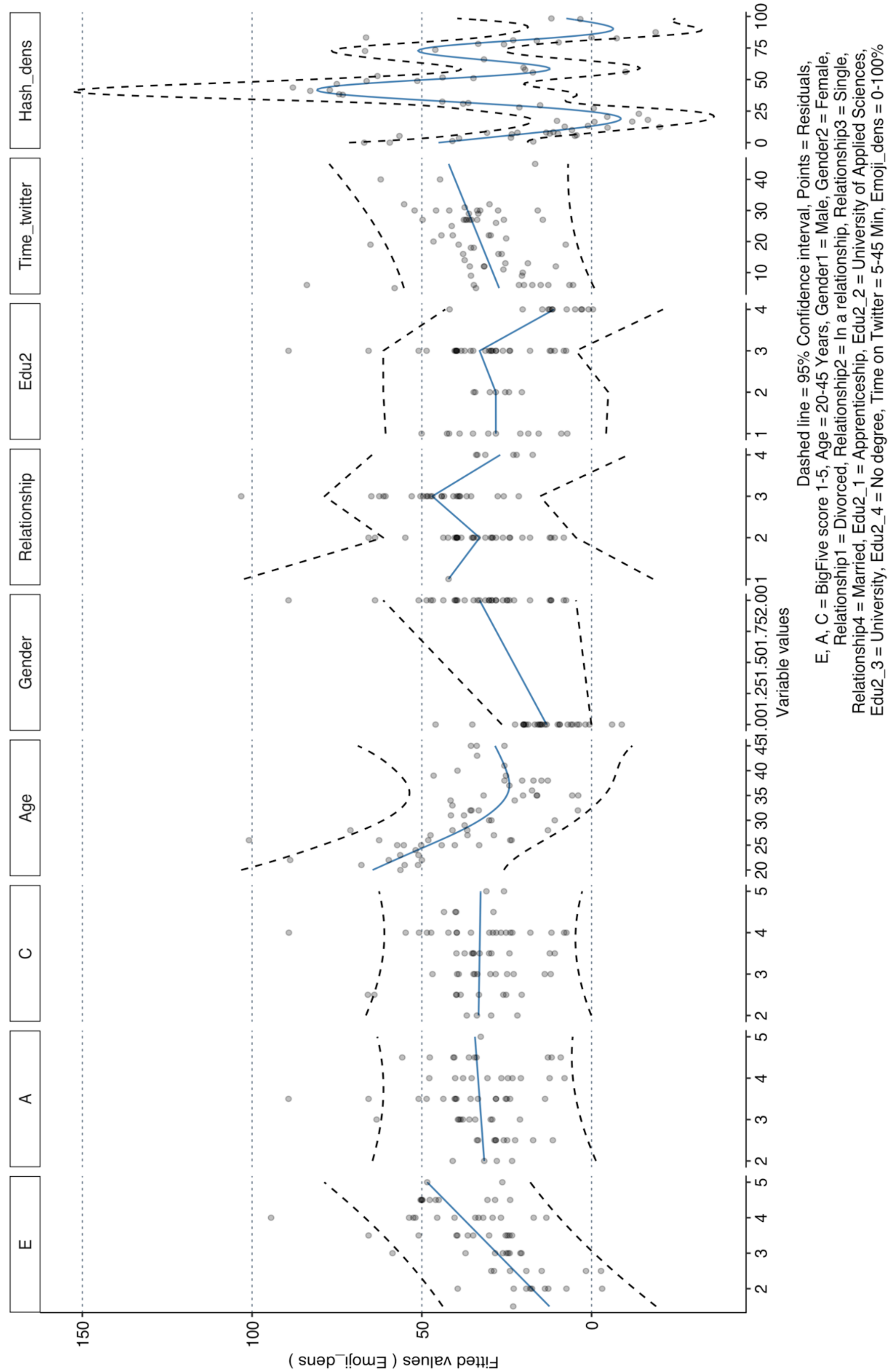

Figure 24. GAM-Plot: Emoji Density ~ Predictors.

According to the results of the GAM (Table 20), hypothesis (2d) was rejected. However, and this is where the GAM really plays at its strengths compared to a linear model, when we look at the non-linear interactions between age and gender in Figure 24, we find that they are significant for both males and females. The first three variables (E, A, C) represent the linear portion of the GAM and the corresponding plots are residual plots like in a simple linear regression model. Further, Figure 24 shows a trend that, albeit non-significant (potentially due to the size and/or nature of the sample), indicates that females use more emojis, which is in alignment with current research (Bamman et al., 2014; Hutchins, 2015, October 14; SwiftKey, 2015, April 21b), and that, in relationship to age, there is a significant drop in the usage of emojis with its trough around 35 years increasing again towards 45. Accordingly, previous findings that showed that age does not play a discriminating role (Hutchins, 2015, October 14) have to be amended for this sample. The relationship is simply more complex when age interacts with gender, which, naturally, it always does. Oleszkiewicz et al. (2017) found age and gender to be significant predictors of the number of emojis used by American Facebook users. While they used a different social medium for a different cultural sample, a common trend seems to emerge here.

Table 20 also reveals that extroversion is a significant linear predictor of emoji density. Current research confirms this finding: extraversion has been linked to emoji use with significant correlations especially with positive emojis (Marengo, Giannotta, & Settanni, 2017). This is reflected in Table 10 above, which contains predominantly positive emojis. However, this finding has to be taken with a grain of salt, since agreeableness and neuroticism have also been shown to predict emoticon/emoji use on Facebook (Oleszkiewicz et al., 2017).

4.8.4 Gender Effects and LIWC Categories

Since gender effects on linguistic variables of the same origin (LIWC) were tested, the models were kept consistent throughout categories to see how the same demographic predictor variables (gender and age) affected the outcome variables (different LIWC word categories). Specifically, a parsimonious linear model with gender and age as predictors was used to test the hypotheses. While the normality assumption is usually not a big problem for regression models, the outcome variables were log-transformed if the respective histogram as well as individual D'Agostino tests of skewness (D'Agostino, 1970)[50] indicated moderate to substantial positive (right) skew to normalize or 're-express' the respective outcome variable (Howell, 2007; Tabachnick & Fidell, 2013; Tukey, 1977).[51] If the outcome variable contained zero-values, a constant *a* was added (`a = 1 - min(variable)`) to resolve `-Inf`-values prior to transformation. The model was fit striking a balance between the adjusted R^2, and the AIC score to avoid overfitting. The models for gender and age effects on linguistic variables take the following general form, model 8:

(8) `lm([log] variable [+a]) ~ gender + age, data = liwc_use)`

> (3a) There will be a significant prediction of positive emotion words (as measured by the percentage of words in the positive emotion word category) by gender.

[50] D'Agostino tests of skewness were run using the **`moments`** R-package (Komsta & Novomestky, 2015).
[51] Using a Box-Cox transformation (Box & Cox, 1964) with Yeo-Johnson (2000) power transformation for negative/zero values was also considered, but since the differences in the explained variance were miniscule, the more straightforward log-transformation was chosen (a constant *a* was added when necessary to compensate non-zero values).

Hypothesis (3a) was tested running model 8 with percentage of positive emotion words (log) as the dependent variable, see model 8.

Table 21: *Linear Model Results: Dependent Variable: Positive Emotion Words*

Fixed effects	β	SE	t	p
Intercept	1.04	0.29	3.61	$.0006^{***}$
Genderfemale	0.39	0.12	3.33	$.002^{**}$
Age	-0.00	0.01	-0.07	$.94^{NS}$
$R^2 = .16; R^2_{adj} = .13$				

Signif. codes: 0.001 '***', 0.01 '**', not significant 'NS'

The overall model was significant, $F(2,59) = 5.68$, $p \leq .01$. Table 17 indicates that while age is not a significant predictor, gender is; thus hypothesis (3a) can be confirmed. The beta-coefficient in Table 21 further indicates that when switching from male to female, the amount of positive emotion words increases ($\beta = 0.39$) confirming the stereotype that women use more words related to emotion.

(3b) There will be a significant prediction of positive feeling words (as measured by the percentage of words in the positive emotion word category) by gender.

Hypothesis (3b) was tested running model 8 (see above) with percentage of positive feeling words (log + *a*) as the dependent variable.

Table 22: *Linear Model Results: Dependent Variable: Positive Feeling Words*

Fixed effects	β	*SE*	*t*	*p*
Intercept	-0.07	0.15	-0.51	$.61^{NS}$
Genderfemale	0.16	0.06	2.70	$.009^{**}$
Age	0.01	0.004	2.01	$.05^{*}$
$R^2 = .17$; $R^2_{adj} = .12$				

Signif. codes: 0.01 '**', 0.05 '*', not significant 'NS'

Again, the overall model was significant, $F(2,59) = 5.07$, $p \leq .01$. Table 22 reveals that both gender and age are significant predictors for the percentage of positive feeling words used by the participants. Switching from male to female increases the log-percentage of positive feeling words by a factor of 0.16, ceteris paribus, while a one year increase in age also increases the log-percentage of positive feeling words very slightly ($\beta = 0.01$), ceteris paribus. Hypothesis (3b) can thus also be confirmed.

Hypotheses (3a) and (3b) are discussed together here since positive feeling words are a subcategory of positive emotion words and the implications of the findings are practically the same. The findings are in alignment with previous research, which showed that woman do in fact use more positive emotion words in English (Kokkos & Tzouramanis, 2014; Mehl & Pennebaker, 2003; Newman et al., 2008; Schwartz et al., 2013; Thomson & Muracher, 2001). Here, these findings can be corroborated for a German Twitter context, which indicates that there do not seem to be inter-language, or cross-cultural differences.

> (3c) There will be a significant prediction of negative emotion words (as measured by the percentage of words in the negative emotion word category) by gender.

Hypothesis (3c) was tested using the template in model 8 with percentage of negative emotion words as the outcome variable. Here, the overall model was not significant, $F(2,59) = 1.41$, $p = .25$, and neither gender nor age turned out to be significant predictors for negative emotions words. Hypothesis (3c) was thus rejected, which means that there does not seem to be a significant prediction in how many (percentage) of negative emotion words are used by gender and age. This is in alignment with previous research (Newman et al., 2008), and thus corroborates and extends findings on gender differences and negative emotion words for a German language setting on Twitter. Since Mehl and Pennebaker (2003) suggest that this could be tied to the exact type of negative emotion words (or a subcategory thereof, such as anger), I also ran model 8 with words related to anger (log anger + a).

Table 23: *Linear Model Results: Dependent Variable: Anger Words*

Fixed effects	β	*SE*	*t*	*p*
Intercept	0.94	0.14	6.63	1.13e-08***
Genderfemale	0.12	0.06	2.05	.05*
Age	0.01	0.004	1.94	.06^{NS}
$R^2 = .17$; $R^2_{adj} = .12$				

Signif. codes: 0.01 ‘**’, 0.05 ‘*’, not significant ‘NS’

The results of the linear model in Table 23 indicate that there is in fact a significant prediction of words related to anger by gender. However, it is again the women, who, on average, use slightly more anger words than the men ($\beta = 0.12$), ceteris paribus. This is at odds with Mehl and Pennebaker’s (2003) suggestion that men use more words related to anger, but corroborates the claim that there is a small, but significant gender difference overall.

(3d) There will be a significant prediction of swear words (as measured by the percentage of words in the swear word category) by gender.

Table 24: *Linear Model Results: Dependent Variable: Swear Words*

Fixed effects	β	*SE*	*t*	*p*
Intercept	0.97	0.15	6.51	1.85e-08***
Genderfemale	0.16	0.06	2.70	.01*
Age	0.01	0.01	1.71	.09NS
$R^2 = .13$; $R^2_{adj} = .10$				

Signif. codes: 0.01 '**', 0.05 '*', not significant 'NS'

Hypothesis (3d) was tested using the template in model 8 with percentage of swear words as the outcome variable. The overall model was significant, $F(2,59) = 4.47$, $p = .02$, (see Table 24). And since gender significantly predicts the percentage of swear words, hypothesis (3d) can also be confirmed. However, switching from male to female, ceteris paribus, the amount of swear words in fact increases ($\beta = 0.16$) very slightly, indicating that German female Twitter users use more swear words than their male counterparts. This is quite different from previous, showing that men use more swear words than women in English (Kokkos & Tzouramanis, 2014; Mulac et al., 1986; Schwartz et al., 2013). However, as we have seen above, the use of swear words could also indicate solidarity, and serve as a positive politeness strategy in the context of agreeableness (Daly et al., 2004; Graziano & Eisenberg, 1997; McLeod, 2011; Pinker, 2008), potentially indicating that women, who score higher on agreeableness, use swear words for this purpose. The increase, when switching from male to female, however is only 0.16% ($\beta = 0.16$), so further research is necessary to substantiate this finding.

(3e) There will be a significant prediction of tentative words (as measured by the percentage of words in the tentative word category (see Appendix C, p. 258)) by gender.

Table 25: *Linear Model Results: Dependent Variable: Tentative Words*

Fixed effects	β	*SE*	*t*	*p*
Intercept	1.76	0.33	5.29	1.88e-06***
Genderfemale	-0.27	0.13	-2.01	.05*
Age	-0.02	0.01	-2.05	.05*
$R^2 = .11$; $R^2_{adj} = .08$				

Signif. codes: 0.01 '**', 0.05 '*', not significant 'NS'

Hypothesis (3e) was tested using the template in model 8 with percentage of tentative words as the outcome variable. For model results, refer to Table 25. For hypothesis (3e), the overall model, $F(2,59) = 3.66$, $p = .03$ was significant, and both gender and age turned out to be significant predictors of the percentage of tentative words (for complete list, see Appendix C, p. 258); thus, this finding confirms hypothesis (3e). While there is a miniscule decrease ($\beta = -0.02$) in tentative words with every year increase in age, ceteris paribus, it is interesting that, when switching from male to female, the percentage of tentative words decreases by a factor of 0.27, ceteris paribus. This is at odds with previous research, which had men and not women associated with lower numbers of tentative emotion words in English (Mehl & Pennebaker, 2003; Mulac et al., 2001; Newman et al., 2008). This finding shows that, in a German language context on Twitter, it is the males who use more words related to tentativeness. Inter-gender differences are conceivable; i.e. German female Twitter users' language is less tentative in nature than that of their male counter parts, if only to a small extent. Further, this finding must be limited to the arena of social media, and Twitter in particular, as we cannot extrapolate that German women follow this usage pattern in general, just like previous research has shown that both genders adopt different language usage patterns depending on the context, such as the workplace, for example (Holmes, 2006). The information gleaned from the data so far suggests that females use Twitter for professional purposes quite a bit, thus extending their 'workplace,' and, potentially, adjusting their language accordingly. While this might be at odds with the higher percentage of swear words used by women, there are two conceivable explanations for this: (1) The time of day and the day of the week could be moderating variables, which are not factored in here, i.e. this might only be true for certain times and days, and (2) the increase in percentage of swear words, when switching from male to female, is very low (β = 0.16). This miniscule difference could thus be attributable to the sample used in this study. A comparative analysis would have to be conducted to confirm this, however. Also, the social compensation hypothesis (Correa et al., 2010; Valkenburg & Peter, 2007) might be at play here as well, leading women to be less tentative on a hybrid genre, such as Twitter, just like they would be in the workplace. Naturally, conclusions can only be based on the context of the genre and the social media environment.

(3f) There will be a significant prediction of words related to social concerns (as measured by the percentage of words in the social concerns category) by gender.

Table 26: *Linear Model Results: Dependent Variable: Social Words*

Fixed effects	β	SE	t	p
Intercept	0.99	0.14	7.09	1.91e-09***
Genderfemale	0.14	0.06	2.47	.02*
Age	0.01	0.00	1.58	.12 NS
$R^2 = .12$; $R^2_{adj} = .09$				

Signif. codes: 0.01 '**', 0.05 '*', not significant 'NS'

Hypothesis (3f) was tested using the template in model 8 with percentage of social words (log + *a*) as the outcome variable. According to the model results in Table 26, hypothesis (3f) can also be confirmed by virtue of the fact that gender turned out to be a significant predictor of the log-percentage of words related to social concerns. In particular, switching from male to female increases the amount by a factor of 0.14, ceteris paribus. The overall model, $F(2,59) = 3.88$, $p = .03$ was significant. This is in alignment with previous research (Newman et al., 2008; Schwartz et al., 2013), which has shown that females in an English language environment, talked more about social concerns and discussed people more than men, for instance. Recall that Schwartz et al. (2013) used Facebook data, Newman et al. (2008) used electronic text samples from an archive, thus covering two distinct genres. Here, the involved vs. informational distinction of language comes to mind; research has suggested that females connect speakers and their audiences at an emotional level (Argamon et al., 2003; Biber, 1989; Tannen, 1982). Words related to social concerns fall within this category.

(3g) There will be a significant prediction of words related to family (as measured by the percentage of words in the family category) by gender.

Hypothesis (3g) was tested using the template in model 8 with percentage of words related to family as the outcome variable. Hypothesis (3g) has to be rejected since the overall model was not significant, $F(2,59) = 1.57$, $p = .21$, and neither gender nor age turned out to be significant predictors for family words. This is interesting since statistically significant results would have made sense in light of hypotheses (3f) and (3h) being confirmed with significant models. Since hypothesis (3g) was not confirmed, no further projections can be made pertaining to who uses more words related to family, males or females.

(3h) There will be a significant prediction of percentage of words related to friends (as measured by the percentage of words in the friends category) by gender.

Table 27: *Linear Model Results: Dependent Variable: Friends Words*

Fixed effects	β	SE	t	p
Intercept	0.87	0.14	6.14	7.55e-08***
Genderfemale	0.15	0.06	2.62	.01*
Age	0.01	0.00	1.97	.05*
$R^2 = .14$; $R^2_{adj} = .11$				

Signif. codes: 0.001 '***', 0.05 '*', not significant 'NS'

Hypothesis (3h) was tested using the template in model 8 with percentage of friends words (log + *a*) as the outcome variable. For model results, refer to Table 27. Hypothesis (3h) can be confirmed as well, with gender and age being significant predictors of the log-percentage of words related to friends and the overall model also being significant, $F(2,59) = 4.81$, $p = .01$. Females do indeed use more words pertaining to friends by a factor of 0.15, ceteris paribus. A one year increase in age yields only a miniscule increase (0.01), ceteris paribus. Arguably, the same conclusions as for hypothesis (3f) can be drawn here; Females use words related to friends more than men, which could be considered more involved versus merely informational with a focus on the propositional content.

(3i) There will be a significant prediction of words related to occupation (as measured by the percentage of words in the occupation word category) by gender.

Table 28: *Linear Model Results: Dependent Variable: Occupation Words*

Fixed effects	β	SE	t	p
Intercept	0.89	0.14	6.25	5.02e-08***
Genderfemale	0.15	0.06	2.62	.01*
Age	0.01	0.00	1.87	.06NS
$R^2 = .14$; $R^2_{adj} = .11$				

Signif. codes: 0.001 '***', 0.05 '*', not significant 'NS'

Hypothesis (3i) was tested using the template in model 8 with percentage of occupation words (log + *a*) as the outcome variable. For hypothesis (3i), the overall model, $F(2,59) = 4.66$, $p = .01$ was significant, see Table 28. Gender turned out to be a significant fixed effect, while age did not; the hypothesis was thus confirmed. Interestingly, females use more words related to occupation than men ($\beta = 0.15$). It has become clear, that gender-specific language differences can be highly context dependent. Newman et al. (2008), for example, found that men talk significantly more about words related to occupation than women in English. However, this difference is potentially genre-specific, and as mentioned above, if Twitter serves as an extended workplace for professionals, these findings confirm previous research (Holmes, 2006). It is thus conceivable that the women in this study used Twitter more for their professional lives, resulting in higher frequencies of words related to occupation, unlike the male participants who used Twitter more for private tweets.

(3j) There will be a significant prediction of words related to job (as measured by the percentage of words in the job category) by gender.

Hypothesis (3j) was tested using the template in model 8 with percentage of words related to job as the outcome variable. Since the overall model was not significant, $F(2,59)$ =

1.88, $p = .16$, and neither gender nor age turned out to be significant predictors for job words, hypothesis (3j) has to be rejected. This hints at a trend in this study which debunks some of the previous findings (at least in part), which usually found that men are more informative, talking more about topics related to their jobs. Considering the results for hypothesis (3i) above, it is conceivable that the insignificance of the model here is attributable to the sample size. If we were to extrapolate, then similar findings could be expected since words from the job-category likely have a strong semantic relationship with words from the occupation category.

(3k) There will be a significant prediction of words related to achievements (as measured by the percentage of words in the achievement category) by gender.

Hypothesis (3k) was tested using the template in model 8 with percentage of words related to achievement (log + a) as the outcome variable. Again, the overall model was not significant, $F(2,59) = 1.81$, $p = .17$, and neither gender nor age turned out to be significant predictors for achievement words. Hypothesis (3k) thus has to be rejected as well. This finding indicates that gender did not play a discriminating role in terms of words related to performance/achievements. It seems as if German Twitter users are in equilibrium when it comes to the use of this word-category. The performance category has previously been predominantly associated with males in an English language context (Coates, 2008; Kiesling, 2007, 2011). As such, this finding further indicates potential disparities between how different genders use language in both English and German.

(3l) There will be a significant prediction of words related to money (as measured by the percentage of words in the money category) by gender.

Table 29: *Linear Model Results: Dependent Variable: Money Words*

Fixed effects	β	*SE*	*t*	*p*
Intercept	1.07	0.13	8.27	1.95e-11***
Genderfemale	0.16	0.05	3.00	.004**
Age	0.01	0.00	1.75	.09^{NS}
$R^2 = .16$; $R^2_{adj} = .13$				

Signif. codes: 0.001 '***', 0.01 '**', not significant 'NS'

Hypothesis (3l) was tested using the template in model 8 with percentage of money words (log + *a*) as the outcome variable. As can be inferred from the model results in Table 29, hypothesis (3l) is confirmed since the overall model was significant, $F(2,59) = 1.81$, $p \leq .01$, and gender turned out to be a significant predictor of the log-percentage of words related to money. However, the situation appears to be reversed in comparison to previous research as it is the females, who use more money words by a factor of 0.16, ceteris paribus, and not the males. This is at odds with previous research, which suggested that men rather than women talk more about money (Kiesling, 1997; Newman et al., 2008). However, these findings are contentious to some extent, since 'money-talks' have been shown to be context dependent, e.g. the workplace versus private life (Holmes, 2006) pertaining to different communities of practice — more on this below (Eckert & McConnell-Ginet, 1992), which result in the egalitarian use of language about, e.g. money, for both males and females. This finding could be taken as another hint that female German Twitter users are using the social medium more for professional purposes than for private conversations or statements.

(3m) There will be a significant prediction of words related to sports (as measured by the percentage of words in the sports category) by gender.

Hypothesis (3m) was tested using the template in model 8 with percentage of words related to sports as the outcome variable. Hypothesis (3m) has to be rejected since the overall model was not significant, $F(2,59) = 2.18$, $p = .12$, and neither gender nor age turned out to be significant predictors for sports words. Again, this finding is surprising, because males have been shown to use more words related to sports in previous research (Coates, 1993; Johnson, 1994; Newman et al., 2008; Schwartz et al., 2013). Again, this could mean that there is no significant relationship between gender and sports words for German Twitter users, indicating that males and females use sports words with equal or similar frequencies.

4.8.5 Summary of Gender Effects and LIWC categories

Table 30: *Summary of LIWC Categories and Gender-Age Regression Results (Betas)*

LIWC category	Gender (ref = male)	Age	R^2_{adj}
Positive emotion words	0.39^{**}	-0.00^{NS}	.13
Positive feeling words	0.16^{**}	0.01^{*}	.12
Negative emotion words	Model not	significant	
Anger words	0.12^{*}	0.01^{NS}	.12
Swear words	0.16^{*}	0.01^{NS}	.10
Tentative words	-0.27^{*}	-0.02^{*}	.08
Social words	0.14^{*}	0.01^{NS}	.09
Family words	Model not	significant	
Friends words	0.15^{*}	0.01^{*}	.11
Occupation words	0.15^{*}	0.01^{NS}	.11
Job words	Model not	significant	
Achievement words	Model not	significant	
Money words	0.16^{**}	0.01^{NS}	.13
Sports	Model not	significant	

Signif. codes: 0.01 ‘**’, 0.05 ‘*’, not significant ‘NS’

As the previous hypothesis tests on gender and LIWC categories and their results show (summarized in Table 30), we are presented with a varied picture of findings. Overall, small but significant gender differences were found just as in previous research in English (cf. Newman et al., 2008). While some of the perceptions of how women use language seem to be confirmed, such as their more involved, emotional content, e.g. positive emotion words, or words related to friends (Aramon, Koppel 2003, Biber, 1989), at the same time, other preconceptions are debunked as we are presented with female participants' language use that is quite apart from what research in an English context has found, such as females' lower use of tentative words, their more frequent use of money words, or words related to occupation (Mehl & Pennebaker, 2003; Newman et al., 2008; Schwartz et al., 2013). These findings warrant further investigation into the specific topics males and females talk about. Twitter certainly qualifies as an environment where its users form a community of practice, better yet communities of practice, depending on the language used, or the hashtag (topic) to denote group membership, for instance. While Twitter is very much a physical space housed in servers, and has the form of a corporation, it does fulfil the requirements of a community of practice as being "a community [that is] defined by social engagement" and that is "defined simultaneously by its membership and by the practice in which that membership engages" (Eckert & McConnell-Ginet, 1992, p. 94). As such, the German Twitter community of practice seems to be different from the US (English) community of practice of social media, as it might influence the way German males and females use language differently, thus, not always conforming to previously assumed notions or perceptions about gendered language use. Twitter seems to contribute to creating an environment, an extended workplace for example, in which females adopt more male language patterns. We can certainly see that, since some of the stereotypes have been confirmed, but others undoubtedly, and maybe somewhat unexpectedly, debunked, the issue is not as easy as tying gender to specific word categories, and use those as clear cut predictors of gender — the language and its speakers matter. It seems that especially the females in this study "draw from a very wide and varied discursive repertoire, ranging from normatively 'feminine' to normatively 'male' (Holmes, 2006, p. 1), as evidenced in their use of normatively male word categories.

Referring to the studies mentioned earlier in which gender prediction algorithms where built and tested with varying accuracy in an English-speaking context, it seems clear now that if the same models/algorithms would have been applied here, the outcomes would have looked quite different for gender prediction in a German context. In this study, the same algorithms would not function well if they were based on the notion that females use more tentative words, for example. This, would ultimately decrease accuracy and in fact, warrants the amendment of existing gender prediction algorithms to a particular language context.

The situation seems to be different for German Twitter users compared to English (mostly US) Twitter users from previous studies, with German females using many of the previously associated male word categories relating to achievement, money, job, and tentativeness.

4.8.6 Gender Effects and Word-Based Measures

(4a) There will not be a significant difference between the lexical diversity of men and women as measured by Carroll's CTTR.

Hypothesis (3a) was tested with a one-way ANCOVA with gender as the categorical predictor, age as continuous covariate, and the participants' CTTR scores as dependent variable (see model 9).

(9) `aov(cttr ~ gender + age, data = diss_data)`

Table 31: *ANCOVA Results for the Effect of Gender and Age on CTTR Scores*

Variable	*df*	*F*	ω^2	*p*
Gender	1	5.54	.07	$.02^{*}$
Age	1	1.93	.02	$.17^{NS}$

Signif. codes: 0.05 '*', not significant 'NS'

Hypothesis (4a) has to be rejected as there are significant differences between males and females and scores for the corrected type-token ratio (see Table 31). Factoring in the rather low omega-squared values, we learn that the differences are miniscule. This finding is different from previous research, which did not find any significant differences between males and females and lexical diversity (Alami et al., 2013). This shows that there are in fact significant gender differences and CTTR scores, with males having slightly higher scores than females. This means that the males, on average, produced more lexically varied texts by using a greater variety of individual lexical items (cf. Table 9 above).

(4b) There will not be a significant difference between the vocabulary richness of men and women as measured by Yule's K.

Hypothesis (4b) was also tested with a one-way ANCOVA with gender as the categorical predictor, age as continuous covariate, and the participants' log-Yule's K scores as dependent variable, see model 10.

(10) `aov(log(yules_k) ~ gender + age, data = diss_data)`

Table 32: *ANCOVA Results for the Effect of Gender and Age on Yule's K Scores*

Variable	*df*	*F*	ω^2	*p*
Gender	1	1.11	.002	$.30^{NS}$
Age	1	0.85	.003	$.17^{NS}$

Signif. codes: 0.05 '*', not significant 'NS'

The ANCOVA results in Table 32 show that there are no significant differences in between genders and their Yule's K-scores, which means that hypothesis (4b) can be confirmed. This finding is in alignment with previous research (Alami et al., 2013). Recall that Yule's K measures the overall difficulty of a text. This finding suggests that, on average, there were no significant differences between males and females and the difficulty of the texts (tweets) they produced in this sample.

Summarizing the results of the hypothesis tests for (4a) and (4b), we can see that the relative difficulty of the tweets did not differ between genders even though male tweets were a little more lexically varied.

> (4c) German tweets will show a more 'oral-like' style despite Twitter being a hybrid, mostly written, genre (as measured by the percentages of the two conjunctions *weil* and *denn* 'because,' the former being used in a more informal genre and the latter almost exclusively being used in formal language (Wegener, 1999)).

To test hypothesis (4c), an unpaired Welch two sample t-test was run with the log-transformed variables (adding 1 as a constant). **Unlike** hypothesized, the results indicate that there are no significant differences between the two conjunctions, $t(114.7) = -1.64$, $p = .11$, thus debunking hypothesis (4c). As a non-parametric alternative, a Wilcoxon rank sum test was run with the untransformed variables, confirming the results, $W = 2188.5$, $p = .17$. Both test results indicate that *weil* ‘because,’ the informal alternative, is used significantly more. Thus, German Twitter users debunk previous findings (Scheffler, 2014; Wegener, 1999). German tweets seem to be in equilibrium pertaining to the formality, with which they are written. That further underscores that fact that Twitter is used both for private and professional purposes by the participants. that German tweets are less formal and thus lean towards the oral end of the written-oral hybrid spectrum. The insights gleaned from the multivariate analysis of percentages of punctuation marks corroborates this finding, as overall, few punctuation marks were used. The use of the period, the only statistically significant outlier, is tied to the Big Five trait agreeableness, and age with higher scores/age indicating higher use of punctuation marks. While the period could be a real sign of formality, the diminishingly small percentages of other punctuation marks together with the statistical significance being tied to age and the character trait known to be more adherent to rules outweigh the indicated level of formality.

CHAPTER FIVE: CONCLUSION

5.1 Conclusion

This study examined the interaction between language, personality, and demographic information, such as gender and age, going beyond the confines of a sample exclusively collected in a campus-lab environment. To that end, I sought to achieve the following objectives: (1) to look into the relationship between personality and linguistic (LIWC) categories, and (2) to investigate the effects gender and age have on these LIWC categories, together with other Twitter-related measures, such as hashtag density, emoji density, and sentiment scores.

It turned out that, for the sample in this study, there were small, but statistically significant gender effects on Big5 BFI-10 scores. In addition, there were small significant effects of the Big Five personality dimensions on BFI-10 scores (see Table 7 above).

The tweet data set comprises 19,772 tweets produced by $N = 62$ participants. In this regard, it turned out that while age was a significant predictor of the number of tweets, gender was not, indicating that both genders use tweets with roughly the same frequencies, or at least with significantly indistinguishable frequencies. The hour of the day, or when the tweet was sent, and age were significant predictors of the number of tweets; gender, however, was not.

While previous studies found significant correlations between Big5 categories and positive emotion words in English (Golbeck, Robles, Edmondson, et al., 2011; Küfner et al., 2010; Mehl, 2006), the data in this study revealed significant correlations of personality dimensions (extraversion, agreeableness, and conscientiousness) and the positive feeling category, supporting research that has found significant correlations for both LIWC word categories, and extraversion and agreeableness (Mairesse et al., 2007; Yarkoni, 2010). Thus conscientiousness was added to this list. Since this has not been established in previous research, potential cross-cultural implications are conceivable, indicating that for German Twitter users, high scores on conscientiousness and a high percentage of positive feeling words are not mutually exclusive, playing into the stereotype that Germans gain happiness from being precise. This could also be attributable to variability in the sample. Further investigation of the issue revealed that both extraversion and conscientiousness significantly predict the percentage of positive feeling words. So does the non-linear interaction of age and gender, i.e. the relationship does not follow a straight line, but rather a curvilinear pattern. Interestingly, however, this is only for female participants. In contrast, the male participants follow a linear pattern. This furnishes new evidence for a U-shape of happiness throughout different ages (Blanchflower & Oswald, 2008; Jeste & Oswald, 2014) for the females, while at the same time showing a linear relationship for males (Charles et al., 2001; López et al., 2013). The non-linear interaction of emoji density and gender also significantly predicted the percentage of positive feeling words, indicating that German females on Twitter use more positive emojis, which, in turn, is mirrored by a higher use of positive feeling words.

Different from previous research (Holtgraves, 2011; Mehl, 2006), there was no significant correlation between agreeableness and swear words. Since there was no significant relationship between agreeable German Twitter users and the use of swear words in the corpus, potential cross-linguistic differences between English and German in terms of language use, and personality, as well as cross-cultural differences are at play. However, the Big5 neuroticism score is significantly and positively correlated to the frequency of anxiety words, supporting previous research in different contexts, such as blogs, and emails (e.g. Gill et al., 2006; Nowson, 2006; Yarkoni, 2010).

The data further support previous lines of research (Oerlemans & Bakker, 2014; Pavot et al., 1990; Salary & Shaieri, 2013), which have converged on the conclusion that the extraversion dimension significantly predicts the sentiment score. The findings at hand thus confirm a significant relationship between extraversion and sentiment scores, and by proxy, happiness or positivity expressed in tweets via their emoji usage in a German language environment. Furthermore, German extroverted Twitter users seem to transfer their offline sociability to social media. Arguably, this is a case of the rich-get-richer hypothesis (Correa et al., 2010; Valkenburg & Peter, 2007), indicating that extroverts gain satisfaction from their activity on Twitter. This was also confirmed by extraversion being a significant predictor for the percentage of tweets that contained at least one emoji.

Neuroticism did not affect the sentiment score significantly, which is counter to previous research (Salary & Shaieri, 2013). German Twitter users' scores on the neuroticism dimension thus do not seem to have any bearing on how they 'feel' online. Overall, however, these findings indicate that there is no homogeneous alignment with prior research, underscoring the variability in different populations and cultures, and the inherent uniqueness of language use in relationship to personality traits and social media behavior. Factoring in gender, I found that German female Twitter users seem to be happier than their male counterparts, as the sentiment score increases significantly for women. While age, in contrast, did not significantly affect the sentiment score, the non-linear interaction of age and gender did turn out to be a significant predictor of the percentage of tweets with emojis for both genders. While previous findings in an English speaking context have found the opposite to be true (Hutchins, 2015, October 14), this finding elucidates the complex interaction of gender and age in relationship to emoji use in a German speaking context, thus contributing to a possible cross-cultural trend between different social media. For example, Oleszkiewicz et al. (2017) found age and gender to be significant predictors of the number of emojis used by American Facebook users.

The data also revealed that gender was neither a significant predictor of hashtag density, i.e. the percentage of tweets with at least one hashtag, nor was it a significant predictor for the type of hashtag (tag vs. commentary), which contradicts current research (Shapp, 2014). The Twitter users in this study use both hashtag types indiscriminately for knowledge management (tags) and self-expression (commentary) (Shapp, 2014), and in doing so, do not show gender enactment through different types of hashtags (2,666 hand-coded: 1,155 male, 1,511 female). The usage of hashtags provides support for a more content-related tweet behavior for both male and female German Twitter users, underscoring the importance of the tweet-language, and the impact it has on how Twitter is used (Hong et al., 2011).

The German participants also used both English and German indiscriminately for both tag and commentary hashtags. This not only highlights German Twitter users' proficiency in English, but also their ambition to participate in Twitter conversations and topics that go beyond the boundaries of German speaking countries. Overall, however, the participants did produce more German hashtags.

The middle position, i.e. a hashtag embedded in the syntax of a tweet, significantly predicted the hashtag type. In agreement with previous research on English (Zappavigna, 2015), German Twitter tag-hashtags can be incorporated into the syntax of a tweet much easier than a longer commentary hashtag, in addition to having a higher propensity for being propagated online (Cunha et al., 2011). As the more canonical positions, the middle and end positions were favored over the beginning position for both tag and commentary hashtags.

In terms of LIWC categories, gender turned out to be a significant predictor for both positive emotion and positive feeling words, as well as words related to social concerns and words related to friends, with females using more words in both categories. Here, the findings are in alignment with previous research (Kokkos & Tzouramanis, 2014; Newman et al., 2008; Schwartz et al., 2013; Thomson & Muracher, 2001). Gender also turned out to be a significant predictor for anger words, swear words, tentative words, occupation words, and words related to money. However, the women used a higher percentage of words in these LIWC categories, except for tentative words, of which they used fewer. While previous research (Mulac et al., 1986; Newman et al., 2008; Schwartz et al., 2013) showed that males use more words in these categories, except for tentative words, of which they use fewer, this finding indicates the opposite. German females on Twitter use words from categories that have been associated with males in the past in an English-speaking environment.

Thus, this study offers a perspective into the differences associated with gender and context-dependent language use, and the uniqueness of culturally dependent language use. It appears as though the females in this study use Twitter for a specific, maybe more professional, purpose, not only joining a community of practice of professional female Twitter users, but also entering a specific context, in which language use is adjusted to the context and different usage patterns are adopted to suit the situation, e.g. 'money talks', i.e. conversations about money and financial issues (Holmes, 2006). This has also been shown in previous research on contextualized male-female language use in a professional setting. For example, Mulac et al. (2000) found crossover language use in male and female managers who occupy the same leadership roles, with the females using less references to emotion, more negations, and more oppositions than males when providing criticism. By extrapolation, this study supports these findings by adding Twitter as a professional context, in which crossover language use happens. Thus, these findings contribute to sociolinguistic research in that they show that while some perceived notions about language use hold true, others do not, indicating more egalitarian language use of German Twitter users. They also underscore the importance of not tying a specific gender to specific linguistic patterns, or particular lexical items by making blanket statements about clear-cut, gendered-language use, thus prompting a re-thinking of this issue.

In terms of word-based measures, the participants in this study produced significant differences between genders and CTTR scores, but no significant differences between gender and Yule's K scores. This indicates that the relative difficulty of lexical items in the tweets did not differ between genders, even though tweets written by males were a little more lexically varied. In terms of formality, the results indicate that *weil* 'because,' the informal alternative to *denn* 'because,' was used significantly more. Thus, German Twitter users corroborate previous findings (Scheffler, 2014; Wegener, 1999), indicating that German tweets are less formal and thus lean towards the oral end of the written-oral hybrid spectrum. This is in alignment with the findings pertaining to punctuation. Overall, few punctuation marks were used. Only the period, a statistically significant outlier, was tied to the Big Five trait agreeableness, and age with higher scores/age indicating higher use of periods. While the period could be a real sign of formality, the diminishingly small percentages of other punctuation marks together with the low statistical significance of age and agreeableness, the character trait known to be more adherent to rules, outweigh the indicated level of formality, thus confirming the more informal nature of German tweets.

In sum, the analyses in this study demonstrate small, but systemic language differences between German Twitter users and English speakers, especially in the US, in similar social media contexts pertaining to gender. By using an online corpus in conjunction with accurate demographic information, and computerized text analysis, this study contributes a more solid empirical footing for the controversial topic of gender-based language differences. Previous studies have, at times, confirmed gender-stereotypes pertaining to language use, and thus contributed to the perpetuation of the perception of these stereotypes, and the notion that they have an underlying iota of truth. By attempting to shed new light on how males and females communicate differently on Twitter in a German language setting, this study called into question several previous findings, as the results here offer an opposing point of view, while at the same time confirming other findings. To reiterate, the analyses at hand merely investigated how men and women use language differently on Twitter, not why and only offers suggestions as to why these differences might exist on Twitter.

5.2 Summary of Important Contributions

For expository purposes, I present the main contributions to existing research in an itemized list, couched in the context of previous findings.

- ➢ A natural sample population with more in-depth and more accurate demographic information, together with a naturalistic language tweet sample, expanded the limitations of previous lab samples, or automatically imputed demographic information.
- ➢ Significant positive correlations of extraversion, and agreeableness with positive feeling words (Mairesse et al., 2007; Yarkoni, 2010) could be confirmed for a German Twitter context, adding conscientiousness to the list, which could indicate potential cross-cultural differences pertaining to personality and language use.

- General additive modeling brought to light a significant non-linear relationship between the interaction of gender and age, and positive feeling words for females, furnishing new evidence for a U-shape of happiness throughout different ages (Blanchflower & Oswald, 2008; Jeste & Oswald, 2014), while at the same time showing a linear relationship for males.
- A significant positive correlation between neuroticism and anxiety words was also confirmed (Gill et al., 2006; Nowson, 2006), while the lack of a significant correlation between agreeableness and swear words contradicts prior findings (e.g. Holtgraves, 2011).
- In relationship to sentiment scores and extraversion, a significant relationship was confirmed, supporting the rich-get-richer hypothesis for extraverts on social media (Correa et al., 2010; Valkenburg & Peter, 2007). In addition, this study found a significant relationship between extraversion and emoji density, again confirming current research (Marengo et al., 2017).
- The absence of a significant negative relationship between sentiment scores and neuroticism in this data set contradicts previous research (Salary & Shaieri, 2013).
- Previous findings that women are happier across cultures (Zweig, 2015) were confirmed here for a German Twitter context, based on sentiment scores, women's more frequent use of emojis, and relatively higher scores on the agreeableness and openness Big Five domains.
- Gender did not turn out to have any bearing on either the frequency of tweets with hashtags, or the number of hashtags in the hand-coded subset, nor on the type of hashtag

used by either gender. This finding is different from previous research (Shapp, 2014), thus contradicting the notion of gender enactment through different types of hashtags.

- ➢ The usage of hashtags furnishes evidence for a more content-related tweet behavior for both male and female German Twitter users, underscoring the importance of the tweet-language, and the impact it has on how Twitter is used (Hong et al., 2011).
- ➢ A more complex significant non-linear interaction of gender and age in relationship to emoji density was shown here for German Twitter users, supplementing previous research (Oleszkiewicz et al., 2017), which has found a linear relationship between these parameters in US Facebook users.
- ➢ Gender turned out to be a significant predictor for both positive emotion and positive feeling words, as well as words related to social concerns and words related to friends, with females using more words in both categories. These findings corroborate previous research (Kokkos & Tzouramanis, 2014; Newman et al., 2008; Schwartz et al., 2013; Thomson & Muracher, 2001).
- ➢ Gender also turned out to be a significant predictor for anger words, swear words, tentative words, occupation words, and words related to money. However, it is the women who used a higher percentage of words in these categories, except for tentative words, of which they used fewer. These findings are quite different from previous research in an English (US) context, which showed that the males use more words in these categories, except for tentative words, of which they use fewer (Mulac et al., 1986; Newman et al., 2008; Schwartz et al., 2013), indicating more egalitarian language use on Twitter.

- The German tweet data in this sample support the notion of Twitter as a hybrid genre, in which German users lean more towards informal language use (Scheffler, 2014).

5.3 Implications

5.3.1 Business (Marketing)

As some of the findings in this study indicate that there are significant differences as to how both genders utilize Twitter as a social medium, the study has implications for (online) businesses, marketing, and not least, Twitter itself. For example, the females in this study used emojis more and with greater variety, which could translate into new business communication strategies incorporating emojis. It is also conceivable that new targeted marketing strategies improve Twitter's standing in Germany, where it ranks 7th behind WhatsApp, Facebook, and Facebook messenger among others (Kemp, 2016), making it more competitive, and a more accessible social medium for both male and female Germans. The different usage patterns of males and females can be extrapolated to other social media, where these findings could translate into gender-specific marketing strategies. To reiterate, the active social media penetration in Germany is only 36% (29 million) (Kemp, 2016), so there is definitely room for growth and improvement in terms of social media saturation.

5.3.2 Theoretical Linguistic Research

While the design and rationale of the study was based on previous research and its findings, the results of this study showed that gender differences are not as clear as they are sometimes made out to be. Proceeding with this background in mind, rather surprising and sometimes unexpected results came to light, underscoring the importance of avoiding blanket statements pertaining to gendered-language and clear cut dichotomous language distribution among both genders. This indicates that sociolinguistic research, which factors in social media contexts thus opening new research avenues is only barely scratching the surface. The current study is thus only a steppingstone for further research revolving around the interaction of language, personality, and demographics in relationship to social media, and the different genres, communities of practice, and levels of formality that come with them, e.g. Reddit vs. Instagram, vs. Snapchat, vs. Facebook, vs. more professional social networks, such as LinkedIn and Indeed. Although the sample size is on the smaller end of the spectrum, the statistical significance of many of the hypothesis tests, allows conservative and careful generalizations regarding the larger German Twitter user base.

5.3.3 Cultural Implications

Language inherently is a reflection of different cultures, and the various speech communities therein. Considering the fact that different cultures, and thus languages, do not only realize and practice communication differently, but also adopt different linguistic patterns, amending their language use when necessary, cross-linguistic differences seem to play a pronounced role here. This is especially true for the findings pertaining to gender as a predictor of the frequencies of LIWC categories. Due to the fact that there are cross-cultural similarities and differences, blanket statements a la "this is what culture A does, and this is what culture B does" should be avoided, just like blanket statements on gendered-language use.

5.4 Limitations

While the Big Five factor model is the gold-standard for personality assessments, it does not factor in crucial distinctions among personality traits. The broader Big Five domains, and the more specific facets are grouped together, while, as Weisberg, DeYoung, and Hirsh (2011) claim, “there is no consensus as to the identity and number of facets within the Big Five” (p. 1). Furthermore, the Big Five cannot explain human behavior and experience adequately, relying on simple, non-contingent, comparative statements (McAdams, 1992). This does not diminish the overall efficacy of the model, but should be considered when planning research involving the Big Five.

As with any software, LIWC has some inherent flaws that need to be mentioned. While it is a very effective tool for linguistic analysis, one major flaw is that it is not able to factor in context. Thus, it cannot be used with an n-gram algorithm to account for neighboring words, for example. In addition, metaphors, ambiguous meanings, and homonyms are also not recognized (Tausczik & Pennebaker, 2010). Like any other tool that is dictionary-based, the accuracy of the analyses depends on the quality of the dictionary (Gill et al., 2006). Based on the English dictionary from 2001, the German LIWC dictionary is now over 16 years old. While there are updates for the English version from 2007 (Pennebaker, Booth, et al., 2007) and 2015 (Pennebaker, Booth, et al., 2015), the new version of the German dictionary is still in the works (Wolf, M., personal communication, November 30, 2016), and thus overdue for an update and re-evaluation.

While it is true that true representative samples do not exist, with the sample size here was sufficiently large, providing good representation of both gender and age in addition to accurate statistical results. Nonetheless, the sample size could have been bigger to reduce variability and increase explained variances, and thus increase statistical power in some of the models. A bigger sample would also further increase the generalizability of statistically significant results. Thus, some of the explained variances are low to moderate, which leaves room for the error term, including variables that potentially have explanatory power, but were not factored in here. In addition, with more data, i.e. a bigger sample population, more complex models would have been possible further revealing complex statistical relationships, while at the same time reducing the risk of overfitting. This way, the likelihood of type I errors (false positives) and type II errors (false negatives) can also be further reduced, and, as a result, the risk of the sample differing substantially from the population is minimized.

The noisy text normalization algorithm could have been more fine-grained. Algorithms of this sort are error-prone, since the regular expressions were written by hand, which potentially introduced the possibility for inaccuracies in the results during the text-cleaning process.

The sentiment scores are based on Novak et al.'s (2015b) sentiment ranking of 751 emojis. However, today there are 2,623 emojis in the current UNICODE set (Unicode.org, 2017), which amounts to a difference of 1,872 emojis, which could not be captured in this study. A more extensive list of sentiment-ranked emojis would thus make the sentiment scores more accurate.

Depending on a user's tweet behavior and their online avidity (up to 3,200 tweets can be collected (Twitter Inc., 2017c)), only a snapshot of their overall Twitter language use within a short period of time was collected. This was pointed out as something to be aware of in previous research (Hoffmann, 2007). For some users, this translated into a micro-corpus encompassing every tweet they ever posted, while for other users this resulted in a corpus encompassing only a fraction of the tweets they wrote on Twitter. In addition, the combination of questionnaires with users' tweets results in a smaller number of tweets overall, compared to studies, which used millions of tweets for their analyses.

Finally, the data collection process and/or the questionnaire could be stream-lined to counter-act participant attrition (here, roughly 62%), and invite more participants to provide accurate Twitter handles, without the fear of giving away all-too personal data.

5.5 Recommendations for Future Research

I recommend that research be expanded on the LIWC categories, which were used in this study, and how they pattern according to gender and age, and potentially other demographic variables, in a different German language and social media context, e.g. LinkedIn to compare female's language use there with the results at hand. The findings in this study could also be supplemented by a confirmatory study with the same parameters but a bigger sample size. Along these lines, a true comparative analysis with the exact same parameters, and participants from a different language background is conceivable as well since existing research provides many useful insights, but often uses either participants from different samples, or investigates different social media, or still other genres of written text production.

While the study at hand relied on natural language data from Twitter, there are multiple other social networking sites, such as Instagram, Facebook, LinkedIn, or Snapchat, that still need to be investigated further, especially for a German context with the Big Five factors in mind, and similar demographic information. These social media offer valuable natural, and relatively easily accessible, language data that have the potential to further linguistic and psychological research alike. Against the backdrop of the findings in this study, I recommend future research consider how women use Twitter specifically, i.e. more for business, private lives, or both.

Furthermore, exploratory cluster analysis (hierarchical, k-means, or fuzzy clustering) has the potential to unearth interesting connections that go beyond gender and age differences, and to offer insights into how different participants cluster together.

In addition, topic modeling, e.g. Latent Dirichlet Allocation, suggests itself for the analysis of hashtags and the tweets themselves. In addition, topic modeling should be used to establish a potential connection between gender, LIWC categories, and the topics that are being talked about in tweets. This way, future research could address the question whether the topic of the tweet or hashtag conveys gender, and if there are statistically significant differences as to which topics are preferred by either gender.

REFERENCES

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin, 97*, 129-133.

Aciman, A., & Rensin, E. (2009). *Twitterature: The world's greatest books retold through Twitter*. London, UK: Penguin.

Alami, M., Sabbah, M., & Iranmanesh, M. (2013). Male-female discourse differences in terms of lexical density. *Journal of Applied Sciences, Engineering, and Technology, 5*(23), 5365-5369.

Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., . . . Arslan, R. (2017). rmarkdown: Dynamic Documents for R. R package version 1.4. Retrieved from https://cran.r-project.org/package=rmarkdown

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycholexical study. *Psychological Monographs* (Vol. 47, (Whole No. 211)).

Amichai-Hamburger, Y., & Ben-Artzi, E. (2000). The relationship between extraversion and neuroticism and the different uses of the Internet. *Computers in Human Behavior, 16*, 441-449.

Amichai-Hamburger, Y., Wainapel, G., & Fox, S. (2002). On the Internet no one knows I'm an introvert: Extroversion, neuroticism, and Internet interaction. *CyberPsychology & Behavior, 5*, 125-128.

APA. (2017). Ethical principles of psychologists and code of conduct. Retrieved March 23, 2017 from http://www.apa.org/ethics/code/index.aspx

Application-programming-interface. (2017). *Wikipedia.* Retrieved April 15, 2017 from https://en.wikipedia.org/wiki/Application_programming_interface

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text, 23*(3), 321-346.

Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2007). Mining the blogosphere: Age, gender, and the varieties of self-expression. *First Monday, 12*(9).

Aries, E., & Fern, L. J. (1983). Close friendship in adulthood: Conversational content between same-sex friends. *Sex Roles, 9*, 1183-1196.

Azar, B. (2000). A web of research: They're fun, they're fast, and they save money, but do web experiments yield quality results? *Monitor on Psychology, 31*, 42-47.

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411-421.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychological Science, 21*(3), 372-374.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics, 18*(2), 135-160.

Barbour, S., & Stevenson, P. (1998). *Variation im Deutschen* (K. Gebel, Trans.). Berlin, New York: de Gruyter.

Baron, N. S. (2008). *Always on: Language in an online and mobile world*. New York, NY: Oxford University Press.

Barrick, M., & Mount, M. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology, 78*(1), 111-118.

Bastos, M. T., Galdini Raimundo, R. L., & Travitzki, R. (2012). Gatekeeping Twitter: Message diffusion in political hashtags. *Media, Culture & Society, 0*(0), 1-11.

Bates, D. M., Maechler, M., & Bolker, B. M. (2015). lme4: Linear mixed-effects models using S4 classes. R package version 1.1-13.

Bauer, T. (2017, June 13). State of social media: Das sind die beliebtesten Plattformen Deutschlands. *OnlineMarketing.de.* Retrieved June 20, 2017 from https://onlinemarketing.de/news/social-media-plattformen-deutschland

Baumgarten, F. (1933). Die Charaktereigenschaften *Beiträge zur Charakter- und Persönlichkeitsforschung*. Bern: A. Franke.

Benet-Martinez, V., & John, O. P. (1998). Los cinco grandes across cultures and ethnic groups: Multitrait method analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*, 729-750.

Berend, N. (2005). Regionale Gebrauchsstandards: Gibt es sie und wie kann man sie beschreiben? In L. M. Eichinger & W. Kallmeyer (Eds.), *Standardvariation: Wie viel Variation verträgt die deutsche Sprache?* Berlin: de Gruyter.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Berman, S., DalleMule, L., Greene, M., & Lucker, J. (2012, September 25). Simpson's paradox: A cautionary tale in advanced analytics. *Significance.* Retrieved March 27, 2017 from https://www.statslife.org.uk/the-statistics-dictionary/2012-simpson-s-paradox-a-cautionary-tale-in-advanced-analytics

Berry, D. (2012). Introduction: Understanding the digital humanities. In D. Berry (Ed.), *Understanding digital humanities* (pp. 1-20). London: Palgrave.

Beurskens, M. (2014). Legal questions of Twitter research. In K. Weller, J. E. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 123-133). New York: Peter Lang.

Biber, D. (1989). A typology of English texts. *Linguistics, 27*(1), 3-44.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, England: Cambridge University Press.

Blanchflower, D. G., & Oswald, A. J. (2008). Is well-being U-shaped over the life cycle? *Social Science & Medicine, 66*(8), 1733–1749.

Boucher, J., & Osgood, C. E. (1969). The pollyhanna hypothesis. *Journal of Verbal Learning and Verbal Behavior, 8*, 1-8.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, (Series B) 26*(2), 211-252.

Bruns, A. (2012). How long is a tweet? Mapping dynamic conversation networks on Twitter using Gawk and Gephi. *Information, Communication & Society, 15*(9), 1323-1351.

Bruns, A., & Burgess, J. E. (2011). *The use of Twitter hashtags in the formation of ad hoc publics.* Paper presented at the 6th European Consortium for Political Research, Reykjavik, Iceland.

Burda Forward. (2015). Social trends - social media. *Burda Forward.* Retrieved June 21, 2017 from https://www.burda-forward.de/fileadmin/customer_files/public_files/downloads/studien/BF_SocialTrends_SocialMedia.pdf

Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011). *Discriminating gender on Twitter.* Paper presented at the Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK.

Burgess, J. E., & Bruns, A. (2015). Easy data, hard data: The politics and pragmatics of Twitter research after the computational turn. In G. Langlois, J. Redden, & G. Elmer (Eds.), *Compromised data: From social media to big data* (pp. 93-111). New York: Bloomsbury Academic.

Business Insider. (2017, May 11). Instagram is estimated to gain lots of users in Germany this year. *Business Insider.* Retrieved June 20, 2017 from http://www.businessinsider.com/instagram-is-estimated-to-gain-lots-of-users-in-germany-this-year-2017-5

Caffo, B. (2015). *Regression models for data science in R* Retrieved from https://leanpub.com/regmods

Carroll, J. B. (1964). *Language and thought*. Englewood Cliffs, NJ: Prentice Hall.

Casey, S. (2017). 2016 Nielsen social media report. *Nielsen.* Retrieved June 21, 2017 from http://www.nielsen.com/us/en/insights/reports/2017/2016-nielsen-social-media-report.html

Cashmore, P. (2009, August 6). Twitter zombies: 24% of tweets created by bots. *Mashable.* Retrieved February 20, 2017 from http://mashable.com/2009/08/06/twitter-bots/ - 1tnyHEvaCuqK

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal Social Psychology, 38*, 476-506.

Cattell, R. B. (1948). The primary personality factors in women compared to those in men. *British Journal of Psychology, 1*, 114-130.

Chamorro-Premuzic, T. (2007). *Personality and romantic relationships* (Vol. Personality and individual differences). Boston: Wiley-Blackwell.

Chang, H. C. (2010). A new perspective on Twitter hashtag use: Diffusion of innovation theory. *Proceedings of the Association for Information Science and Technology, 47*(1), 1-4.

Charles, S. T., Reynolds, C. A., & Gatz, M. (2001). Age-related differences and change in positive and negative affect over 23 years. *Journal of Personal Social Psychology, 80*(1), 136–151.

Cheng, N., Chandramouli, R., & Subbalakshmi, K. P. (2011). Author gender identification from text. *Digital Investigation, 8*(1), 78-88.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing, 9*(6), 1-14.

Chung, C., & Pennebaker, J. W. (2007). The psychological function of function words. In K. Fiedler (Ed.), *Social Communication* (pp. 343-359). New York: Psychology Press.

Churches, O., Nicholls, M., Thiessen, M., Kohler, M., & Keage, H. (2014). Emoticons in mind: An event-related potential study. *Social Neuroscience, 9*(2), 196-202.

Clark, A. (2003). *Pre-processing very noisy text.* Paper presented at the Workshop on Shallow Processing of Large Corpora. Corpus Linguistics., Lancaster, UK (pp. 12-22).

Clark, E., & Araki, K. (2011). Text normalization in social media: Progress, problems and applications for a pre-processing system of casual English. *Procedia - Social and Behavioral Sciences, 27*, 2-11.
Clark, M. (2016, June 26). Generalized additive models. Retrieved April 17, 2017 from https://m-clark.github.io/generalized-additive-models/
Cleeton, G. U., & Knight, F. B. (1924). Validity of character judgments based on external criteria. *Journal of Applied Psychology, 8*(2), 215-231.
Coates, J. (1989). Gossip revisited: Language in all-female groups. In J. Coates & D. Cameron (Eds.), *Women in their speech communities* (pp. 94-121). London: Longman.
Coates, J. (1993). *Women, men, and language* (2nd ed.). London: Longman.
Coates, J. (1997). One-at-a-time: The organization of men's talk. In S. Jonson & U. H. Meinhof (Eds.), *Language and Masculinity* (pp. 107 - 129). Cambridge, MA: Blackwell.
Coates, J. (2008). *Men talk*. Malden, MA: Blackwell Publishing.
Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*(12), 1304-1312.
Cordes, G. (1983). *Handbuch zur niederdeutschen Sprach- und Literaturwissenschaft*. Berlin: Erich Schmidt Verlag.
Correa, T., Hinsley, A. W., & de Zúñiga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior, 26*(2), 247-253.
Costa, P., & McCrae, R. (1992). *Revised NEO personality inventory and NEO five factor professional manual*. Odessa, FL: Psychological Assessment Resources.
Costa, P., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322-331.
Côté, S., & Moskowitz, D. S. (1998). On the dynamic covariation between interpersonal behavior and affect: Prediction from neuroticism, extraversion, and agreeableness. *Journal of Personality and Social Psychology, 75*(4), 1032-1046.
Crawford, M. (1995). *Talking difference*. London: SAGE.
Cunha, E., Magno, G., Comarela, G., Almeida, V., Gonçalves, M. A., & Benevenuto, F. (2011). *Analyzing the dynamic evolution of hashtags on Twitter: A language-based approach.* Paper presented at the Workshop on Language in Social Media, Portland, OR.
Cunha, E., Magno, G., Gonçalves, M. A., & Benevenuto, F. (2012). *A gender based study of tagging behavior in Twitter.* Paper presented at the 23rd ACM conference on hypertext and social media, Milwaukee, WI.
D'Arcy, A., & Young, T. M. (2012). Ethics and social media: Implications for sociolinguistics in the networked public1 Ethics and social media: Implications for sociolinguistics in the networked public. *Journal of Sociolinguistics, 16*(4), 532-546.
D'Agostino, R. B. (1970). Transformation to normality of the null distribution of G1. *Biometrika, 57*(3), 679-681.
Daly, N., Holmes, J., Newton, J., & Stubbe, M. (2004). Expletives as solidarity signals in FTAs on the factory floor. *Journal of Pragmatics, 36*, 945-964.
Danet, B. (2001). *Cyberpla@y: Communicating online*. London: Berg.
Darling, W. M., Paul, M. J., & Song, F. (2012). *Unsupervised part-of-speech tagging in noisy and esoteric domains with a syntactic-semantic bayesian hmm.* Paper presented at the EACL Workshop on Semantic Analysis in Social Media, Avignon, France.

Davis, M., & Edberg, P. (2016). Unicode technical report #51. *Unicode Technical Reports.* Retrieved March 26, 2017 from http://unicode.org/reports/tr51/
Deitrick, W., Miller, Z., Valyou, B., Dickinson, B., Munson, T., & Hu, W. (2012). Gender identification on Twitter using the modified balanced winnow. *Communication and Networks, 4*(3), 189-195.
Derks, D., Bos, A., & Grumpkow, J. v. (2007). Emoticons and social interaction on the internet: The importance of social context. *Computers in Human Behavior, 23*, 842-849.
DeWall, N., Buffardi, L., Bonser, I., & Campbell, K. (2011). Narcissim and implicit attention seeking: Evidence from linguistic analyses of social networking and online presentation. *Personality and Individual Differences, 51*, 57-62.
Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*(1), 417.
Dodsworth, R. (2013). Retreat from the southern vowel shift in Raleigh, NC: Social factors. *University of Pennsylvania Working Papers in Linguistics, 19*(2), 29-40.
Doyle, G. (2014). *Mapping dialectal variation by querying social media.* Paper presented at the 14th Conference of the European Chapter of the Association for Computational Linguistics.
Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory (10503293), 20*(3), 249-268.
Duggan, M., & Brenner, J. (2013). The demographics of social media users - 2012. *Pew Internet and American Life Project*.
Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school*. New York: Teach. Coll. Press.
Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics, 12*, 453-476.
Eckert, P. (2011a). Gender and sociolinguistic variation. In J. Coates & P. Pichler (Eds.), *Language and Gender: A reader* (pp. 57 - 70). Malden, MA: Wiley-Blackwell.
Eckert, P. (2011b). Language and power in the preadolescent heterosexual market. *American Speech, 86*(1), 85-97.
Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology, 41*, 87-100.
Eckert, P., & McConnell-Ginet, S. (1992). Communities of practice: Where language, gender, and power all live. In K. Hall (Ed.), *Locating power: Proceedings from the 2nd Berkeley Women and Language Conference* (pp. 89-99). Berkely: BWLG.
Eckert, P., & McConnell-Ginet, S. (1995). Constructing meaning, constructing selves: Snapshots of language, gender, and class from Belten High. In K. Hall & M. Bucholtz (Eds.), *Gender articulated: Arrangements of language and the socially constructed self* (pp. 469-507). New York: Routledge.
Einspänner, J., Dang-Anh, M., & Thimm, C. (2014). Computer-assisted content analysis of Twitter data. In K. Weller, J. E. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 97-108). New York: Peter Lang.
Eisenstein, J. (2013). *What to do about bad language on the internet.* Paper presented at the NAACL-HLT, Atlanta, GA.
Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2010). A latent variable model for geographic lexical variation. *EMNLP '10 Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1277-1287.

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in social media. *Plos One, 9*(11), 1 - 13.

Eleta, I., & Golbeck, J. (2014). Multilingual use of Twitter: Social networks at the language frontier. *Computers in Human Behavior, 41*(Dec), 424-432.

Emogi Research Team. (2015, September 14). 2015 Emoji report. Retrieved October 10, 2016 from http://emogi.com/documents/Emoji_Report_2015.pdf

Emogi Research Team. (2016, November 16). 2016 Emoji report. Retrieved April 25, 2017 from http://cdn.emogi.com/docs/reports/2016_emoji_report.pdf

Eurostat. (2016). Digital economy and society statistics - households and individuals. *Eurostat.* Retrieved March 21, 2017 from http://ec.europa.eu/eurostat/statistics-explained/index.php/Information_society_statistics_-_households_and_individuals -Internet_usage

Evans, V. (2015, November 18). Sign of our times: Why emoji can be even more powerful than words. *The Conversation.* Retrieved January 17, 2017 from https://theconversation.com/signs-of-our-times-why-emoji-can-be-even-more-powerful-than-words-50893

Evans, V. (2017). *The emoji code: The linguistics behind smiley faces and scaredy cats*. New York, NY: Picador.

Facebook. (2017). Company info: Stats. *Facebook Newsroom.* Retrieved March 4, 2017 from http://newsroom.fb.com/company-info/

Fahlmann, S. (1982). Original bboard thread in which :-) was proposed. Retrieved September 23, 2017 from https://www.cs.cmu.edu/~sef/Orig-Smiley.htm

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25*(5), 1-54.

Feingold, A. (1994). Gender differences in personality: A meta analysis. *Psychological Bulletin, 116*, 429-456.

Field, A., Miles, J., & Field, Z. (2013). *Discovering statistics using R*. London: Sage Publications.

Filippova, K. (2012). User demographics and language in an implicit social network. *EMNLP-CoNLL '12: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1,478-471,488.

Fink, C., Kopecky, J., & Morawski, M. (2012). Inferring gender from the content of tweets: A region specific example, 459-462. Retrieved from AAAI Publications, Sixth International AAAI Conference on Weblogs and Social Media website: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4644

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal Social Psychology, 44*, 329-344.

Fitzpatrick, T. B. (1975). Soleil et peau. *Journal de Médecine Esthétique, 2*, 33-34.

Friedrichs, M. (2009). Jung, männlich und gebildet: So sieht die deutsche Twitter-Gemeinde aus [blog]. Retrieved June 12, 2017 from https://www.basicthinking.de/blog/2009/04/02/jung-maennlich-und-gebildet-so-sieht-die-deutsche-twitter-gemeinde-aus/

Frost, J. (2015, September 3). The danger of overfitting regression models [blog]. *The Minitab Blog.* Retrieved April 11, 2017 from http://blog.minitab.com/blog/adventures-in-statistics-2/the-danger-of-overfitting-regression-models

Fuhrhop, N., & Barghorn, R. (2012). Prinzipien der Wortschreibung im Deutschen in Englischen am Beispiel der Schreibdiphthonge und der Doppelkonsonanten. In L. Gunkel & G. Zifonun (Eds.), *Deutsch im Sprachvergleich: Grammatische Kontraste und Konvergenzen* (pp. 135-159). Berlin, Boston: De Gruyter.

Fullwood, C., Quinn, S., Chen-Wilson, J., Chadwick, D., & Reynolds, K. (2015). Put on a smiley face: Textspeak and personality perception. *CyberPsychology, Behavior, and Social Networking, 18*(3), 147-151.

Gelman, A. (2004, December 10). Against parsimony [blog]. *Statistical Modeling, Causal Inference, and Social Science.* Retrieved April 11, 2017 from http://andrewgelman.com/2004/12/10/against_parsimo/

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Gentry, J. (2016). twitteR: R based Twitter client (R package version 1.1.9). Retrieved from https://cran.r-project.org/package=twitteR

German Tagsets. (2017). STTS Tag Table. Retrieved April 24, 2017 from http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/GermanTagsets.html

Gill, A. J., Oberlander, J., & Austin, E. (2006). Rating e-mail personality at zero acquaintance. *Personality and Individual Differences, 40*(3), 497-507.

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). *Predicting personality from Twitter.* Paper presented at the IEEE International Conference on Privacy, Security Risk, and Trust, and IEEE International Conference on Social Computing, Boston, MA.

Golbeck, J., Robles, C., & Turner, K. (2011). *Predicting personality with social media.* Paper presented at the CHI '11 Extended Abstracts on Human Factors in Computing Systems, Vancouver, BC, Canada.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of Personality and Social Psychology*. Beverly Hills, CA: Sage.

Goldberg, L. R. (1992). The development of markers for the Big Five factor structure. *Psychological Assessment, 4*, 26-42.

Goodwin, M. H. (1988). Cooperation and competition across girls' play activities. In T. Dundas & S. Fisher (Eds.), *Gender and discourse: The power of talk* (pp. 55-94). Norwood, NJ: Ablex.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality, 37*, 504-528.

Gosling, S. D., Sandy, C. J., & Potter, J. (2010). Personalities of self-identified dog people and cat people. *Anthrozoos: A Multidisciplinary Journal of The Interactions of People & Animals, 23*, 213-222.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*, 93-104.

Gottschalk, L. A., & Gleser, G. C. (1969). *The measurement of psychological states through the content analysis of verbal behavior*. Berkeley, CA: University of California Press.

Gottschalk, L. A., Stein, M. K., & Shapiro, D. H. (1997). The application of computerized content analysis of speech to the diagnostic process in a psychiatric outpatient clinic. *Journal of Clinical Psychology, 53*, 427-441.

Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 795-824).

Greenwood, S., Perrin, A., & Duggan, M. (2016, November 2016). Social media update 2016 [report]. *Pew Research Center.* Retrieved March 26, 2017 from http://www.pewinternet.org/2016/11/11/social-media-update-2016/

Grefenstette, G., Qu, Y., Evans, D., & Shanahan, J. (2008). Validating the coverage of lexical resources for affect analysis and automatically classifying new words along semantic axes. In J. Shanahan, Y. Qu, & J. Wiebe (Eds.), *Computing attitude and affect in text: Theory and applications* (pp. 93-107). Dordrecht, The Netherlands: Springer.

Hall, K. (1995). Lip service on the fantasy lines. In K. Hall & M. Bucholtz (Eds.), *Gender articulated: Language and the socially constructed self* (pp. 183-215). New York: Routledge.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: E. Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. London: E. Arnold.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Hastie, T. J., & Tibshirani, R. J. (1986). Generalized additive models. *Statistical Science, 1*(3), 297-310.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman and Hall.

Hermida, A. (2014). Twitter as an ambient news network. In K. Weller, J. E. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 359-372). New York: Peter Lang.

Herring, S. C. (2008). Language and the internet. In W. Donsbach (Ed.), *The concise encyclopedia of communication* (pp. 2640-2645). Oxford, UK: Wiley-Blackwell.

Herring, S. C. (2010). Web content analysis: Expanding the paradigm. In J. Hunsinger, M. Allen, & L. Klastrup (Eds.), *The international handbook of internet research* (pp. 233-249). Berlin, DE: Springer Verlag.

Herring, S. C., & Androutsopoulos, J. (2015). Computer mediated discourse 2.0. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed., pp. 127-151). Chichester, UK: John Wiley & Sons.

Herring, S. C., & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics, 10*(4), 439-459.

Herring, S. C., & Zelenkauskaite, A. (2008). Gendered typography: Abbreviation and insertion in Italian iTV SMS. In J. F. Siegel, T. C. Nagel, A. Laurente-Lapole, & J. Auger (Eds.), *IUPWL7: Gender in langauge: Classic questions, new contexts* (pp. 73-92). Bloomington, IN: IULC Publications.

Hess, C. W., Ritchie, K. P., & Landry, R. G. (1984). The type-token ratio and vocabulary performance. *Psychological Reports, 55*(1), 51-57.

Hilker, C., & Raake, S. (2010). Mit weniger mehr erreichen *Bankmagazin* (Vol. 05.10, pp. 26-28). Wiesbaden: Springer Gabler.

Hill, K. (2012, August 9). The invasion of the Twitter bots. *Forbes.* Retrieved February 20, 2017 from http://www.forbes.com/sites/kashmirhill/2012/08/09/the-invasion-of-the-twitter-bots/ - 1c2b48a93273

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality, 43*(3), 524-527.

Hodgkinson, G., & Ford, J. (2008). *International Review of Industrial and Organizational Psychology*.

Hoffmann, S. (2007). Processing Internet-derived text: Creating a corpus of Usenet messages. *Literary and Linguistic Computing, 22*(2), 151-165.

Hogenboom, A., Bal, D., Frasincar, F., Bal, M., de Jong, F., & Kaymak, U. (2015). Exploiting emoticons in polarity classification of text. *Journal of Web Engineering, 14*(1-2), 22-40.

Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (2015). *Probability and statistical inference* (9th ed.). Boston: Pearson.

Holmes, J. (1984). Women's language: A functional approach. *General Linguistics, 24*, 149-178.

Holmes, J. (1995). *Women, men and politeness*. New York: Longman.

Holmes, J. (2001). Gender, politeness, and stereotypes *An introduction to sociolinguistics*. Essex: Pearson Education Limited.

Holmes, J. (2006). *Gendered talk at work*. Malden, MA: Blackwell.

Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of Research in Personality, 45*, 92-99.

Honeycutt, C., & Herring, S. C. (2009). *Beyond microblogging: Conversation and collaboration via Twitter.* Paper presented at the Forty-second Hawai'i International Conference on System Sciences, Los Alamitos, CA.

Hong, L., Convertino, G., & Chi, E. H. (2011). *Language matters in Twitter: A large scale study.* Paper presented at the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain.

Houstin, K. (2013). *Shady characters: The secret life of punctuation, symbols, and other typological marks*. New York: W. W. Norton & Company.

Howell, D. C. (2007). *Statistical methods for psychology* (6th ed.). Belmont, CA: Thomson Wadsworth.

Huang, J., Thornton, K. M., & Efthimiadis, E. N. (2010). *Conversational tagging in Twitter.* Paper presented at the 21st ACM conference on hypertext and hypermedia, Toronto, ON, Canada.

Hudson, M., Nicolas, S., Howser, M., Lipsett, K., Robinson, I., Laura, P., . . . Friedman, D. (2015). Examining how gender and emoticons influence Facebook jealousy. *CyberPsychology, Behavior, and Social Networking, 18*(2), 87-92.

Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior, 28*(2), 561-569.

Hutchins, B. (2015, October 14). The emoji infographic: Stats to back up your obsession [blog]. Retrieved August 31, 2016 from https://www.meltwater.com/blog/the-emoji-infographic-stats-to-back-up-your-obsession/

Iacobelli, F., Gill, A., Nowson, S., & Oberlander, J. (2011). Large Scale Personality Classification of Bloggers. In S. D'Mello, A. Grasser, B. Schuller, & J. C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 568-577): Springer-Verlag GmbH.

Isaac, M. (2017, September 26). Twitter to test doubling tweet length to 280 characters. *The New York Times.* Retrieved October 10, 2017 from https://www.nytimes.com/2017/09/26/technology/twitter-280-characters.html

Java, A., Song, X., Finin, T., & Tseng, B. (2007). *Why we twitter: Understanding microblogging usage and communities.* Paper presented at the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, San Jose, CA.

Jeste, D. V., & Oswald, A. J. (2014). Individual and societal wisdom: Explaining the paradox of human aging and high well-being. *Psychiatry, 77*(4), 317–330.

Johannson, T. (2011). Hail the impossible: p-values, evidence, and likelihood. *Scandinavian Journal of Psychology, 52*(2), 113-125.

John, O. P. (1990). The Big Five factor taxonomy: Dimensions of personality in the natrual language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66-100). New York: Guilford Press.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *Eur. J. Pers, 2*, 171-205.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The "Big Five" inventory: Versions 4a and 5b*. Retrieved from Berkeley:

John, O. P., & Srivastava, S. (2001). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research*. New York: Guilford.

Johnson, S. (1994). A game of two halves? On men, football, and gossip. *Journal of Gender Studies, 3*, 145-154.

Johnstone, B. (2010). Language and place. In R. Mesthrie & W. Wolfram (Eds.), *Cambridge handbook of sociolinguistics*. Cambridge: Cambridge Univeristy Press.

Jost, J., West, T., & Gosling, S. D. (2009). Personality and ideology as determinants of candidate preferences and Obama conversion in the 2008 presidential electioin. *DuBois Review: Social Science Research on Race, 6*(1), 103-124.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Kamps, H. J. (2015, May 25). Who are Twitter's verified users? *Medium.com.* Retrieved February 24, 2017 from https://medium.com/@Haje/who-are-twitter-s-verified-users-af976fc1b032 - .flwkwww7d

Kaufmann, M., & Kalita, J. (2010). *Syntactic normalization of Twitter messages.* Paper presented at the ICON-2010: 8th International Conference on Natural Language Processing, Kharagpur, India.

Kaye, L., Wall, H., & Malone, S. (2016, April 13). What your emojis say about you. *The Conversation.* Retrieved April 13, 2016 from https://theconversation.com/what-your-emojis-say-about-you-57523

Kehoe, A., & Gee, A. (2011). Social tagging: A new perspective on textual 'aboutness'. In P. Rayson, S. Hoffman, & G. Leech (Eds.), *Studies in variation, contacts, and change in English* (Vol. 6). Helsinki: Research Unit for Variation, Contacts, and Change in English.

Kelly, R., & Watts, L. (2015). Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design.* from http://opus.bath.ac.uk/46780/

Kemp, S. (2016). Digital in 2016 [slideshare]. *We Are Social Singapore.* Retrieved March 23, 2017 from https://www.slideshare.net/wearesocialsg/digital-in-2016

Kiesling, S. F. (1997). Power and the language of men. In S. Johnson & U. H. Meinhof (Eds.), *Language and Masculinity* (pp. 65-85). Oxford: Blackwell.
Kiesling, S. F. (2007). Men, masculinities, and language. *Language and Linguistics Compass, 1*(6), 653-673.
Kiesling, S. F. (2011). Playing the straight man: Displaying and maintaining male heterosexuality in discourse. In J. Coates & P. Pichler (Eds.), *Language and gender: A reader* (pp. 275 - 285). Malden, MA: Wiley-Blackwell.
King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Sciene, 30*(3), 666-687.
Klages, L. (1926). *The Science of Character*. London: Allen & Unwin.
Kokkos, A., & Tzouramanis, T. (2014). A robust gender inference model for online social networks and its application to LinkedIn and Twitter. *First Monday, 19*(9), 1-12.
Komsta, L., & Novomestky, F. (2015). Moments: Moments, cumulants, skewness, kurtosis, and related tests (R package version 0.14). Retrieved September 7, 2017 from https://cran.r-project.org/package=moments
König, W. (2005). *dtv-Atlas Deutsche Sprache*. München: Deutscher Taschenbuch Verlag GmbH.
Koppel, M., Argamon, S., & Shimoni, A. (2003). Automatically categorizing written texts by author gender. *Literary and Linguistic Computing, 17*, 101-108.
Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience – the dangers of double dipping. *Nature Neuroscience, 12*(5), 535-540.
Krohn, F. B. (2004). A generational approach to using emoticons as nonverbal communication. *Journal of Technical Writing & Communication, 34*(4), 321-328.
Küfner, A. C. P., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality, 44*, 427-435.
Kurath, H. (1949). *A word geography of the eastern United States*: University of Michigan Press.
Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.
Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2*, 205-254.
Labov, W., Ash, S., & Boberg, C. (2008). *The atlas of North American English: Phonetics, phonology, and sound change*: De Gruyter.
Lakoff, R. (1975). *Language and woman's place*. New York, NY: Harper & Row.
Lameli, A. (2004). *Standard und Substandard: Regionalismen im diachronen Längsschnitt*. Stuttgart: Franz Steiner Verlag.
Larsen, K. (2015, July 30). GAM: The predictive modeling silver bullet [blog]. *MultiThreaded.* Retrieved December 9, 2016 from http://multithreaded.stitchfix.com/assets/files/gam.pdf
Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., . . . van Alstyne, M. (2009). Life in the network: The coming age of computational social science. *Science, 323*(5915), 721-723.
Lee, J. (2009). Is that an emoticon in 1862. *The New York Times.* Retrieved September 23, 2017 from https://cityroom.blogs.nytimes.com/2009/01/19/hfo-emoticon/
Lee, K. (2016, April 27). The biggest social media science study: What 4.8 million tweets say about the best time to tweet [blog]. *Buffer Social.* Retrieved June 12, 2017 from https://blog.bufferapp.com/best-time-to-tweet-research

Lehrer, A. (1974). *Semantic fields and lexical structure*. London: North Holland.
Leppänen, S., Pitkänen-Huhta, A., Piirainen-Marsh, A., Nikula, T., & Peuronen, S. (2009). Young People's Translocal New Media Uses: A Multiperspective Analysis Of Language Choice And Heteroglossia. *Journal of Computer-Mediated Communication, 14*(4), 1080-1107.
Levinson, S. C. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press.
Lin, Y., Margolin, D., Keegan, B., Baronchelli, A., & Lazer, D. (2013). *# Bigbirds never die: Understanding social dynamics of emergent hashtag.* Paper presented at the 7th International AAAI Conference on Weblogs ans Social Media, Boston, MA.
LIWC. (2016, November 2016). How it works. Retrieved November 23, 2016 from http://liwc.wpengine.com/how-it-works/
López, U. B. F., Møller, V., & Sousa-Poza, A. (2013). How does subjective well-being evolve with age? A literature review. *Journal of Population Ageing, 6*(3), 227–246.
Luginbühl, M. (2003). Streiten im Chat. *Linguistic Online, 15*, 70-87. Retrieved from http://www.linguistik-online.de/15_03/luginbuchl.pdf
Lutz, C. (1986). Emotion, thought, and estrangement: Emotion as a cultural category. *Cultural Anthropology, 1*(3), 287-309.
Lutz, C. (1990). Engendered emotion: Gender, power, and the rhetoric of emotional control in American discourse. In L. A. Lughod & C. Lutz (Eds.), *Language and the politics of emotion* (pp. 69-91). Cambridge: Cambridge University Press.
Ma, Z., Sun, A., & Cong, G. (2012). *Will this #hashtag be popular tomorrow?* Paper presented at the 35th international ACM SIGIR conference on research and development in information retrieval, Portland, OR.
Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., & Smith, A. (2013). Teens, social media, and privacy. *Pew Internet and American Life Project.* Retrieved November 14
Maireder, A., & Ausserhofer, J. (2014). Political discourses on Twitter: Networking topics, objects, and people. In K. Weller, J. E. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 305-318). New York: Peter Lang.
Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457-500.
Maltz, D. N., & Borker, R. A. (1982). A cultural approach to male and female miscommunication. In J. Gumperz (Ed.), *Language and identity* (pp. 195 - 216). Cambridge, MA: Cambridge University Press.
Marengo, D., Giannotta, F., & Settanni, M. (2017). Assessing personality using emoji: An exploratory study. *Personality and Individual Differences, 112*(Supplement C), 74-78.
Markham, A., & Buchanan, E. (2012). Ethical decision-making and internet reserach: Recommendations from the AoIR Working Committee (Version 2.0). *AoIR.* Retrieved April 10, 2017 from https://aoir.org/reports/ethics2.pdf
Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data*. London: Lawrence Erlbaum Associates, Inc., Publishers.
McAdams, D. P. (1992). The Five Factor model in personality: A critical appraisal. *Journal of Personality, 60*(2), 329-361.
McCandless, M., & Sanford, M. (2013a). cldr: Language identifier based on CLD library (R package version 1.1.0). Retrieved from ftp://cran.r-project.org/pub/R/src/contrib/Archive/cldr

McCandless, M., & Sanford, M. (2013b). Package cldr (version 1.1.0). Retrieved April 28, 2017 from http://www2.uaem.mx/r-mirror/web/packages/cldr/cldr.pdf
McCrae, R., & Costa, P. (1990). *Personality in adulthood: A five-factor theory perspective*. New York: Guilford Press.
McCrae, R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175-215.
McCrae, R. R. (1989). Why I Advocate the Five-Factor Model: Joint Factor Analyses of the NEO-PI with Other Instruments. In D. Buss & N. Cantor (Eds.), *Personality Psychology* (pp. 237-245): Springer US.
McDougall, W. (1932). Of the words character and personality. *Character Personality, 1*, 3-16.
McGill, A. (2016, May 9). Why white people don't use white emoji. *The Atlantic.* Retrieved March 26, 2017 from https://www.theatlantic.com/politics/archive/2016/05/white-people-dont-use-white-emoji/481695/
McLeod, L. (2011). Swearing in the 'Tradie' environment as a tool for solidarity. *Griffith Working Papers in Pragmatics and Intercultural Communication, 4*(1/2), 1-10.
McMillan, J. R., Clifton, A. K., McGrath, D., & Gale, W. S. (1977). Women's language: Uncertainty or interpersonal sensitivity and emotionality? *Sex Roles, 3*, 545-559.
Mehl, M. R. (2006). Quantitative text analysis. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 141-156). Washington, DC: American Psychological Association.
Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality & Social Psychology, 84*, 857-870.
Messina, C. (2007, August 23). "how!do!you!feel!about!using!#!(pound)!for!groups.!As!in!#barcamp! [msg]?"! [tweet]. *Twitter.* Retrieved March 31, 2017 from https://twitter.com/chrismessina/status/223115412
Messina, C. (2007, August 25). Groups for Twitter - or a proposal for Twitter tag channels [blog]. *Factory Joe.* Retrieved November 13, 2015 from http://factoryjoe.com/blog/2007/08/25/groups-for-twitter-or-a-proposal-for-twitter-tag-channels/
Messina, C. (2007, October 22). Twitter hashtags for emergency coordination and disaster relief [blog]. *FactoryJoe.* Retrieved November 13, 2015 from https://factoryjoe.com/2007/10/22/twitter-hashtags-for-emergency-coordination-and-disaster-relief/
Michalke, M. (2017a). koRpus: An R package for text analysis (R package version 0.10-2). Retrieved February 21, 2017 from https://reaktanz.de/?c=hacking&s=koRpus
Michalke, M. (2017b). koRpus: An R Package for Text Analysis (Version 0.10-2). from https://reaktanz.de/?c=hacking&s=koRpus
Miller, Z., Dickinson, B., & Hu, W. (2012). Gender prediction on Twitter using stream algorithms with n-gram character features. *International Journal of Intelligence Science, 2*(24), 143-148.
Miranda-García, A., & Calle-Martín, J. (2005). Yule's characteristic K revisited. *Language Resources and Evaluation, 39*(4), 287-294.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., . . . Sczesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods, 46*(3), 841-871.
Mowbray, M. (2014). Automated Twitter accounts. In K. Weller, J. E. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 183 - 194). New York: Peter Lang.
Mroczek, D. K., & Kolarz, C. M. (1998). The effect of age on positive and negative affect: A developmental perspective on happiness. *Journal of Personality and Social Psychology, 75*(5), 1333-1349.
Mukherjee, A., & Bing, L. (2010). Improving gender classification of blog authors. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 207-217.
Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research, 27*, 121-152.
Mulac, A., Lundell, T. L., & Bradac, J. J. (1986). Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect. *Communication Monographs, 53*, 115-129.
Mulac, A., Seibold, D. R., & Farris, J. L. (2000). Female and male managers' and professionals' criticism giving: Differences in language use and effects. *Journal of Language and Social Psychology, 19*(4), 389-415.
Mulac, A., Studley, L. B., & Blau, S. (1990). The gender-linked effect in primary and secondary students' impromptu essays. *Sex Roles, 23*, 439-469.
Murphy, D. (2014, April 13). 44 Percent of Twitter accounts have never tweeted. *PCMag.* Retrieved March 15, 2017 from http://www.pcmag.com/article2/0,2817,2456489,00.asp
Mywordismybond[Def.1]. (2017). *Urban Dictionary.* Retrieved April 15, 2017 from http://www.urbandictionary.com/define.php?term=my%20word%20is%20my%20bond
Naaman, M., Boase, J., & Lai, C. H. (2010). *Is it really about me? Message content in social awareness streams.* Paper presented at the Conference on Computer Supported Cooperative Work, Savannah, GA.
Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from Generalized Linear Mixed-effects models. *Methods in Ecology and Evolution, 4*, 133-142.
Neuman, G., Wagner, S., & Christiansen, N. (1999). The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management, 24*(1), 28.
Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes, 45*(3), 211-236.
Nitins, T., & Burgess, J. E. (2014). Twitter, brands, and user engagement. In K. Weller, J. E. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 293-304). New York: Peter Lang.
Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015a). Emoji sentiment ranking v1.0. Retrieved April 29, 2017 from http://kt.ijs.si/data/Emoji_sentiment_ranking/
Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015b). Sentiment of emojis. *Plos One, 10*(12), 1-22.
Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2016). Emoji sentiment map. Retrieved April 29, 2017 from http://kt.ijs.si/data/Emoji_sentiment_ranking/emojimap.html

Nowak, P. (2016, January 2). Twitter turns profit, but it still resists change. *The National.* Retrieved March 25, 2017 from http://www.thenational.ae/business/technology/twitter-turns-a-profit-but-it-still-resists-change

Nowson, S. (2006). *The language of weblogs: A study of genre and individual differences*. Unpublished doctoral dissertation. University of Edinburgh, Edinburgh, UK.

Nowson, S., Oberlander, J., & Gill, A. (2005). Weblogs, genres, and individual differences. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, 1666-1671.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity, 41*, 673-690.

O'Brien, T., & DeLongis, A. (1996). The interactional context of problem-, emotion-, and relationship-focused coping: The role of the big five personality factors. *Journal of Personality, 64*(4), 775-813.

OECD. (2015). Digital government [report]. *OECD.* Retrieved March 28, 2017 from http://www.keepeek.com/Digital-Asset-Management/oecd/governance/government-at-a-glance-2015_gov_glance-2015-en - page147

OECD. (2015, November 15). Soziale Netzwerke und der Bildungsstand ihrer User: Deutschland entgegem dem Trend http://bit.ly/1Od8pbR #Neuland [tweet]. Retrieved March 28, 2017 from https://twitter.com/OECDStatistik/status/665569403283152902/photo/1?ref_src=twsrc%5Etfw&ref_url=http%3A%2F%2Fwww.bento.de%2Fgadgets%2Fsocial-media-bildung-studie-149542%2F

Oerlemans, W. G. M., & Bakker, A. B. (2014). Why extraverts are happier: A day reconstruction study. *Journal of Research in Personality, 50*, 11-22.

Oleszkiewicz, A., Karwowski, M., Pisanski, K., Sorokowski, P., Sobrado, B., & Sorokowska, A. (2017). Who uses emoticons? Data from 86702 Facebook users. *Personality and Individual Differences, 119*(Supplement C), 289-295.

Ong, E., Ang, R., Ho, J., Lim, J., Goh, D., Lee, C. S., & Chua, A. (2011). Narcissim, extraversion, and adolescents' self-presentation on Facebook. *Personality and Individual Differences, 50*, 180-185.

Page, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication, 6*(2), 181-201.

Park, J., Barash, V., Fink, C., & Cha, M. (2013). Emoticon style: Interpreting differences in emoticons across cultures. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, 466-475.

Pasta, D. J. (2009). *Learning when to be discrete: Continuous vs. categorical predictors*. Paper presented at the ICON Clinical Research, San Francisco, CA. Retrieved from http://support.sas.com/resources/papers/proceedings09/248-2009.pdf

Pavalanathan, U., & Eisenstein, J. (2016). More emojis, less :) The competition for paralinguistic function in microblog writing. *First Monday, 21*(7-11), Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/6879.

Pavot, W., Diener, E., & Fujita, F. (1990). Extraversion and happiness. *Personality and Individual Differences, 11*(2), 1299-1306.

Peersman, C., Daelemans, W., & van Vaerenbergh, L. (2011). Predicting age and gender in online social networks. *SMUC '11: Proceedings of the Third International Workshop on Search and Mining User-Generated Content*, 37-44.

Penn Treebank Tag Set. (2017). Penn treebank II tag set. Retrieved April 24, 2017 from http://www.clips.ua.ac.be/pages/mbsp-tags

Pennebaker, J. W., Booth, R., Boyd, R. L., & Francis, M. E. (2015). Linguistic inquiry and word count: Operator's manual. Retrieved April 23, 2017 from https://liwc.wpengine.com/wp-content/uploads/2015/11/LIWC2015_OperatorManual.pdf

Pennebaker, J. W., Booth, R., & Francis, M. (2007). *Linguistic inquiry and word count*. Mahwah, NJ: Erlbaum.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: Univeristy of Texas at Austin.

Pennebaker, J. W., Chung, C., Ireland, M., Gonzales, A., & Booth, R., J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion, 10*, 601-626.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6), 1296-1312.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natrual language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577.

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology, 85*(2), 291-301.

Perrine, R., & Osbourne, H. (1998). Personality characteristics of dog and cat persons. *Anthrozoos: A Multidisciplinary Journal of The Interactions of People & Animals, 11*(1), 33-40.

Peterka-Bonetta, J. (2017a). Emojis [GitHub repository]. Retrieved April 29, 2017 from https://github.com/today-is-a-good-day/emojis

Peterka-Bonetta, J. (2017b). Emojis analysis in R [blog]. *Opiate for the Masses.* Retrieved April 29, 2017 from http://opiateforthemass.es/articles/emoji-analysis/ - disqus_thread

Pfeiffer, T. (2009). Deutsche Twitteranalyse [website: defunct]. *Webevangelist.* Retrieved June 12, 2017 from http://tomro.se/deutsche-twitteranalyse

Pinker, S. (2008). Freedom's curse. *The Atlantic Monthly, 302,* 28-29.

Popov, V., Kosinski, M., Stillwell, D., & Kielczewski, B. (2017). Applymagicsauce. Retrieved November 16, 2017 from https://applymagicsauce.com/

Poushter, J. (2016, February 22). Smartphone ownership and internet usage continues to climb in emerging economies [report]. *Pew Research Center.* Retrieved March 22, 2017 from http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/

Qui, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality, 46*, 710-718.

R Development Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rammstedt, B., & John, O. P. (2005). Short version of the Big Five Inventory (BFI-K): Development and validation of the Big Five inventory in English and German. *Journal of Research in Personality, 41*, 203-212.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five inventory in English and German. *Journal of Research in Personality, 41*, 203-212.
Rammstedt, B., Kemper, C., Klein, M. C., Beierlein, C., & Kovaleva, A. (2012). *Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI-10)*. Mannheim: GESIS.
Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). *Classifying latent user attributes in Twitter.* Paper presented at the 2nd international workshop on search and mining user-generated contents.
Rawlings, D., & Ciancarelli, V. (1997). Music preference and the five-factor model of the NEO Personality Inventory. *Psychology of Music, 25*(5), 120.
Raymond, M. (2010, April 14, April 14, 2016). How tweet it is! Library acquires entire Twitter archive [blog]. *Library of Congress Blog.* Retrieved April 10, 2016 from https://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/
Reddy, S., & Stanford, J. (2015). *A web application for automated dialect analysis.* Paper presented at the NAACL 2015, Denver, CO.
Rehbein, I., & Ruppendorfer, J. (2013). *Investigating orality in speech, writing, and in between*. Paper presented at the Corpus Linguistics 2013, Lancaster, UK.
Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*(6), 1236-1256.
Reyes, A., Rosso, P., & Veale, T. (2013). A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation, 47*(1), 239-268.
Romero, D. M., Meeder, B., & Kleinberg, J. (2011). *Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter.* Paper presented at the International World Wide Web conference, Hyderabad, India.
Rossi, L., & Magnani, M. (2012). *Conversation practices and network structure in Twitter.* Paper presented at the Sitxth international AAAI Conference on Weblogs and Social Media, Dublin, Ireland.
Rothenberg, M. (2013a). Emojitracker: realtime emoji use on Twitter. Retrieved April 28, 2017 from http://www.emojitracker.com/
Rothenberg, M. (2013b). How I built Emojitracker. *Medium.* Retrieved April 28, 2017 from https://medium.com/@mroth/how-i-built-emojitracker-179cfd8238ac
RStudio Team. (2017). RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/
Rudin, M. (2011). From Hemingway to Twitterature: The short and shorter of it. *The Journal of Electronic Publishing, 14*(2).
Russ, B. (2012). *Examining large-scale regional variation through online geotagged corpora*. Paper presented at the Annual Meeting of the American Dialect Society.
Russell, M. A. (2013). *Mining the social web: Data mining Facebook, Twitter, LinkedIn, Google +, Github, and more*. Sebastopol, CA: O'Reilly Media, Inc.
Salary, S., & Shaieri, M. R. (2013). Study of the relationship between happiness and dimensions of psychosis, neurosis, and personality extraversion. *Procedia - Social and Behavioral Sciences, 84*, 1143-1148.
Salgado, J. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment, 10*(1&2), 117-125.

Scheffler, T. (2014). *A German Twitter snapshot.* Paper presented at the International Conference on Language Resources and Evaluation, Reykjavik, Iceland.
Scheffler, T., & Kyba, C. C. M. (2016). *Measuring social jetlag in Twitter data.* Paper presented at the Tenth International AAAI Conference on Web and Social Media (ICWSM2016), Cologne, Germany.
Schifanella, R., Juan, P. d., Tetreault, J. R., & Cao, L. (2016). Detecting sarcasm in multimodal social platforms. *CoRR, abs/1608.02289*.
Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006). Effects of age and gender on blogging. *Proceedings of the AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs Conference*, 27-29.
Schmid, H. (1994). *Probabilistic Part-of-Speech Tagging Using Decision Trees.* Paper presented at the International Conference on New Methods in Language Processing, Manchester, UK.
Schmid, H. (1995). *Improvements in Part-of-Speech Tagging with an Application to German.* Paper presented at the ACL SIGDAT-Workshop., Dublin, Ireland.
Schmid, H. U. (2012). *Bairisch: Das Wichtigste in Kürze*. München: C. H. Beck.
Schnoebelen, T. (2012). Do you smile with your nose? Stylistic variation in Twitter emoticons. *University of Pennsylvania Working Papers in Linguistics, 18*(2), 115-125.
Schupp, J., & Gerlitz, J. (2008). Das BFI-S: Big Five Inventory-SOEP. In A. Glöckner-Rist (Ed.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 12.00*. Bonn: GESIS.
Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., . . . Ungar, L. H. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *Plos One, 8*(9), e73791.
Seargeant, P., & Tagg, C. (Eds.). (2014). *The language of social media: Identity and community on the Internet*. New York: Palgrave MacMillan.
Selfhout, M., Burk, W., Branje, S., Denissen, J., van Aken, M., & Meeus, W. (2010). Emerging late adolescent friendship networks and Big Five personality traits: A social network approach. *Journal of Personality, 78*(2), 509-538.
Semiocast. (2010). Only thiry percent of tweets are from the U.S.; U.S., Japan, Brazil top 3 Twitter nations. Retrieved November 13 from http://semiocast.com/en/publications/2010_03_31_only_thirty_percent_of_tweets_are_from_the_us.
Semiocast. (2010, February 24). Half of messages on Twitter are not in English: Japanese is the second most used language. Retrieved November 13, 2015 from https://semiocast.com/downloads/Semiocast_Half_of_messages_on_Twitter_are_not_in_English_20100224.pdf
Seward, Z. M. (2014, August 11). Twitter admits that as many as 23 million of its active users are automated. *Quartz Media.* Retrieved February 20, 2017 from https://qz.com/248063/twitter-admits-that-as-many-as-23-million-of-its-active-users-are-actually-bots/
Shapp, A. (2014). *Variation in the use of Twitter hashtags*. Department of Linguistics. New York University. Retrieved from https://www.nyu.edu/projects/shapp/Shapp_QP2_Hashtags_Final.pdf

Shaver, P., & Brennan, K. (1992). Attachment styles and the Big Five personality traits: Their connections with each other and with romantic relationship outcomes. *Personality and Social Psychology Bulletin, 18*(5), 536.
Sidarenka, U., Scheffler, T., & Stede, M. (2013). *Rule-based normalization of German Twitter messages.* Paper presented at the International Conference of the German Society for Computational Linguistics and Language Technology, GSCL, Darmstadt, Germany.
Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. Sebastopol, CA: O'Reilly.
Smith, A., & Brewer, J. (2012). *Twitter use 2012*. Retrieved from Technical Report:
Smith, K. (2016, May 17). 44 Twitter statistics for 2016 [blog]. Retrieved March 20, 2017 from https://www.brandwatch.com/blog/44-twitter-stats-2016/
Soziale Medien. (2016, March 21). Twitter nennt erstmals Nutzerzahlen für Deutschland. *Zeit Online.* Retrieved October 19, 2016 from http://www.zeit.de/digital/2016-03/soziale-medien-twitter-nutzerzahlen-deutschland
Stark, L., & Crawford, K. (2015). The conservatism of Emoji: Work, affect, and communication. *Social Media + Society*, 1-11.
Statista. (2016). Most-used languages on Twitter as of September 2013. Retrieved November 21, 2016 from https://www.statista.com/statistics/267129/most-used-languages-on-twitter/
Statista. (2017). Durchschnittliches Alter Geschiedener zum Zeitpunkt der Scheidung in Deutschland von 2000 bis 2015. *Statista.* Retrieved June 20, 2017 from https://de.statista.com/statistik/daten/studie/453647/umfrage/durchschnittsalter-bei-ehescheidung-in-deutschland/
Stieglitz, S., & Krüger, N. (2014). Public enterprise-related communication and its impact. In K. Weller, J. E. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 281-292). New York: Peter Lang.
Stone, A. A., Schwartz, J. E., & Broderick, J. E. (2010). A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Science U S A., 107*(22), 9985–9990.
Suárez-Colmenares, F. (2017). Emojis.csv [GitHub repository - now defunct]. Retrieved April 23, 2017 from https://raw.githubusercontent.com/felipesua/sampleTexts/master/emojis.csv
Surveycircle. (2017, March 15). Got an empty data set? Retrieved March 15, 2017 from https://www.surveycircle.com/en/
SwiftKey. (2015, April 21a). Most-used emoji revealed: Americans love skulls, Brazilians love cats, the French love hearts [blog]. *SwiftKey.* Retrieved March 31, 2017 from https://blog.swiftkey.com/americans-love-skulls-brazilians-love-cats-swiftkey-emoji-meanings-report/
SwiftKey. (2015, April 21b). SwiftKey emoji report. Retrieved March 10, 2016 from https://blog.swiftkey.com/americans-love-skulls-brazilians-love-cats-swiftkey-emoji-meanings-report/
Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
Tagliamonte, S. (2006a). *Analising sociolinguistic variation*. Cambridge, MA: Cambridge University Press.

Tagliamonte, S. (2006b). *Analysing sociolinguistic variation*. New York: Cambridge University Press.
Tagliamonte, S. (2014). Situating media influence in sociolinguistic context. *Journal of Sociolinguistics, 18*(2), 223-232.
Tagliamonte, S., & D'Arcy, A. (2007). Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change, 19*(02), 199-217.
Tanaka-Ishii, K., & Aihara, S. (2015). Computational constancy measures of texts—Yule's K and R´enyi's Entropy. *Computational Linguistics, 41*(3), 481-502.
Tannen, D. (1982). Oral and literate strategies in spoken and written narratives. *Language, 58*(1), 1-21.
Tannen, D. (1990a). Gender differences in conversational coherence: Physical alignment and topical cohesion. In B. Dorval (Ed.), *Conversational coherence and its development* (pp. 167-206). Norwood, NJ: Ablex.
Tannen, D. (1990b). *You just don't understand: Women and men in conversation*. New York, NY: Harper Collins.
Tannen, D. (1991). *You just don't understand: Women and men in conversation*. New York: Harper Collins.
Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24-54.
Thayer, A., Evans, M. B., McBride, A. A., Queen, M., & Spyridakis, J. H. (2010). I, pronoun: A study of formality in online content. *Journal of Technical Writing and Communication, 40*(4), 447-458.
The Digital Policy Council. (2016, January 23). *Prnewswire.com.* Retrieved March 19, 2017 from http://www.prnewswire.com/news-releases/world-leaders-on-twitter--adoption-stagnates-even-as-follower-base-explodes-300208802.html
Thomson, R., & Muracher, T. (2001). Predicting gender from electronic discourse. *British Journal of Social Psychology, 40*, 193-208.
Tippmann, S. (2015). Programming tools: Adventures with R. *Nature, 517*(7532), 109-110.
Tong, S., & Koller, D. (2002). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Reserach, 2*, 45-66.
TreeTagger. (2017). TreeTagger - a part-of-speech tagger for many languages. Retrieved April 10, 2017 from http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/
Tseliga, T. (2007). "It's all Greeklish to me!" Linguistic and sociocultural perspectives on Roman-alphabeted Greek in asynchronous computer-mediated communication. In S. C. Herring & B. Danet (Eds.), *The multilingual internet: Language, culture, and communication online* (pp. 116-142). New York, NY: Oxford University Press.
Tsur, O., & Rappoport, A. (2012). *What's in a hashtag? Content based prediction of the spread of ideas in microblogging communities*. Paper presented at the Fifth ACM international conference on Web search and data mining, Seattle, WA.
Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Election forcasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review, 29*(4), 402-418.

Tupes, E., & Christal, R. (1957). Personality traits related to effectiveness of junior and senior Air Force officers. *USAF Personnel Training Res. Cent. Rep. No.*, 57-125.
Tupes, E., & Christal, R. (1961). Recurrent personality factors based on trait ratings. *USAF ASD Tech. Rep.*, 61-97.
Twenge, J. M., Konrath, S., Foster, J. D., Campbell, W. K., & Bushman, B. J. (2008). A cross-temporal meta-analysis of the Narcissitic Personality Inventory. *Journal of Personality, 76*, 875-902.
Twitter Data. (2016, March 21). Today may be all about 🎂🎉, but the most-used emoji in Twitter's history is 😂. It's appeared on Twitter 14.5 billion times. #LoveTwitter [tweet]. Retrieved from https://twitter.com/TwitterData/status/711950817478234113
Twitter Engineering. (2011, June 30). 200 million tweets per day [blog]. *Twitter.* Retrieved March 20, 2017 from https://blog.twitter.com/2011/200-million-tweets-per-day
Twitter Inc. (2011, March 14). #numbers [blog]. *Twitter.* Retrieved February 9, 2017 from https://blog.twitter.com/2011/numbers
Twitter Inc. (2017a). About verified accounts. *Twitter Support.* Retrieved March 30, 2017 from https://support.twitter.com/articles/119135
Twitter Inc. (2017b). Developer agreement and policy. Retrieved April 28, 2017 from https://dev.twitter.com/overview/terms/agreement-and-policy
Twitter Inc. (2017c). GET statuses/sample. *Twitter.* Retrieved January 17, 2017 from https://dev.twitter.com/streaming/reference/get/statuses/sample
Twitter Inc. (2017d). Terms of service. Retrieved April 28, 2017 from https://twitter.com/tos?lang=en.
Twitter Inc. (2017e). Twitter privacy policy. Retrieved April 28, 2017 from https://twitter.com/privacy?lang=en
Twitter Inc. (2017f). Twitter usage/Company facts. *Twitter.* Retrieved March 15, 2017 from https://about.twitter.com/company
Twitter Inc. (2017g). Using hashtags on Twitter. *Twitter Support.* Retrieved January 17, 2017 from https://support.twitter.com/articles/49309
Twittimer. (2017). Post whenever you want. Retrieved February 24, 2017 from https://twittimer.com/
Unicode.org. (2017). Full emoji data v5.0. Retrieved November 16, 2017 from https://unicode.org/emoji/charts/full-emoji-list.html
Valenzuela, S., Halperna, D., & Katz, J. E. (2014). Social network sites, marriage well-being and divorce: Survey and state-level evidence from the United States. *Computers in Human Behavior, 36*, 94 - 101.
Valkenburg, P. M., & Peter, J. (2007). Preadolescents' and adolescents' online communication and their closeness to friends. *Developmental Psychology, 43*, 267-277.
Vaux, B., & Golder, S. (2003). Harvard dialect survey. from https://www4.uwm.edu/FLL/linguistics/dialect/maps.html
Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*(5), 1202-1216.
Votruba, M. (n.d.). The 17th century emoji. *Slovak Studies Program.*
Wagemakers, A. (2015, August 3). There is a possibility that the quality of Twitter's users is deteriorating. *Business Insider.* Retrieved March 25, 2017 from http://www.businessinsider.com/twitter-monthly-active-users-2015-7?r=UK&IR=T

Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician, 36*(1), 46-48.
Waldman, K. (2016, July 19). Where does “Your Word Is Your Bond” come from, and why did Melania steal it? [blog]. *Slate.* Retrieved April 20, 2017 from http://www.slate.com/blogs/lexicon_valley/2016/07/19/your_word_is_your_bond_history_and_origins_from_matthew_to_hip_hop.html
Wall, H., Kaye, L., & Malone, S. (2016). An exploration of psychological factors on emoticon usage and implications for judgement accuracy. *Computers in Human Behavior, 62*, 70-78.
Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). *A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle.* Paper presented at the ACL 2012 System Demonstrations, Jeju Island, Korea.
Wardrop, R. L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician, 49*(1), 24-28.
WDR. (2016). Abitur 2016: Mehr Mädchen als Jungen bei den Abi-Abschlüssen. Retrieved June 14, 2017 from http://www1.wdr.de/nachrichten/maedchen-jungen-abitur-100.html
Weerkamp, W., Carter, S., & Tsagkias, M. (2011). *How people use Twitter in different languages.* Paper presented at the 3rd International Conference on Web Science, Koblenz, Germany.
Wegener, H. (1999). Syntaxwandel und Degrammatikalisierung im heutigen Deutsch? Noch einmal zu weil-Verbzweit. *Deutsche Sprache, 27*(1), 3-26.
Weintraub, W. (1989). *Verbal behavior in everyday life*. New York, NY: Springer.
Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology, 2*, 1-11.
Weller, K., Bruns, A., Burgess, J. E., Mahrt, M., & Puschmann, C. (Eds.). (2014). *Twitter and society*. New York: Peter Lang.
Werry, C. C. (1996). Linguistic and interactional features of internet relay chat. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (pp. 47-63). Philadelphia, PA: John Benjamins.
Westermann. (2017). Wörter merken - Wörter mit Doppelvokalen. *Kapiert.de.* Retrieved May 3, 2017 from https://www.kapiert.de/deutsch/klasse-7-8/rechtschreibung/woerter-merken/woerter-merken-woerter-mit-doppelvokalen/
Whelan, S., & Davies, G. (2006). Profiling consumers of own brands and national brands using human personality. *Journal of Retailing and Consumer Services, 13*(6), 393-402.
Whitlock, T. (2017). Emoji unicode tables. *apps.timwhitlock.info.* Retrieved April 29, 2017 from http://apps.timwhitlock.info/emoji/tables/unicode
Wickham, H. (2014). Tidy data. *Journal of Statistical Software, 59*(10), 1-23.
Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis, and Generalized Additive Modeling. *Journal of Language Evolution, 1*(1), 7-18.
Wolf, A. (2000). Emotional expression online: Gender differences in emoticon use. *CyberPsychology & Behavior, 3*(5), 827-833.
Wolf, M., Horn, A. B., Mehl, M. R., Haug, S., Pennebaker, J. W., & Kordy, H. (2008). Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count. *Diagnostica, 54*(2), 85-98.
Wolfram, W., & Schilling-Estes, N. (1997). *Hoi toide on the outer banks: The story of the Ocracoke Brogue*. Chapel Hill, NC: The University of North Carolina Press.

Wolfram, W., & Schilling-Estes, N. (2006). *American English: Dialects and Variation*. Malden, MA: Blackwell Publishers.
Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65*(1), 95-114.
Wood, S. N. (2006). *Generalized additive models: An introduction with R* (Vol. 66). Boca Raton, FL: Chapman and Hall/CRC.
Word[Def.3]. (2017). *Urban Dictionary.* Retrieved April 15, 2017 from http://www.urbandictionary.com/define.php?term=Word&utm_source=search-action
Yang, L., Sun, T., Zhang, M., & Ei, Q. (2012). *We know what @You #Tag: Does the dual role affect hashtag adoption.* Paper presented at the 21st International Conference on World Wide Web, New York, NY.
Yarkoni, T. (2010). Personality in 100,00 words: A large scale analysis of personality and word use among bloggers. *Journal of Research in Personality, 44*, 363-373.
Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87*(4), 954-959.
Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society, 13*(5), 788-806.
Zappavigna, M. (2015). Searchable talk: the linguistic functions of hashtags. *Social Semiotics, 25*(3), 274-291.
Zhao, H., & Seibert, S. (2006). The Big Five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology, 91*(2), 259-271.
Zhu, L. (2010, January 20). When do most people tweet? At the end of the week. [blog]. *Hubspot.* Retrieved June 12, 2017 from https://blog.hubspot.com/blog/tabid/6307/bid/5500/When-Do-Most-People-Tweet-At-the-End-of-the-Week.aspx
Zou, H., & Hastie, T. (2005). Regulatization and variable selection via the elastic net. *Journal of the Royal Statistical Society, 67*, 301-320.
Zweig, J. S. (2015). Are women happier than men? Evidence from the Gallup world poll. *Journal of Happiness Studies, 16*(2), 515-541.

APPENDICES

Appendix A: Questionnaire – German Version

Teil I – Einverständniserklärung/Informed consent

Informed consent – German version
Project Title:
Interactions[3] – language, demographics, and personality; an in-depth analysis of German tweets

The German version was translated by the PI, a German native speaker.

Willkommen zur Umfrage und vielen Dank für deine/Ihre Zeit!

Einleitung

Im Rahmen eines Promotionsvorhabens an der Ball State University möchten wir Menschen, die Twitter aktiv nutzen bitten einen kurzen Online-Fragebogen auszufüllen.

Das Ziel der vorliegenden Studie ist eine Verbindung zwischen den linguistischen Merkmalen deutscher Twitter Nutzer, demographischen Informationen, und der Persönlichkeit der Nutzer herzustellen und nachzuweisen.

Der Zeitaufwand beträgt ungefähr **fünf** Minuten!

Ausschlusskriterien

Teilnehmen können Menschen, die Twitter als soziales medium aktiv nutzen (mit einem öffentlichen Profil) und zwischen 18 und 45 Jahren alt sind.

Denk/denken Sie daran, es gibt keine (!) falschen Antworten; all Antworten sind wertvoll für diese Umfrage.

Bonus

Als Dankeschön für deine/Ihre Teilnahme and meiner Studie, hast du/haben Sie die Möglichkeit einen von drei 20 EUR Amazon Gutschein zu gewinnen. Die Gewinner werden automatisch und blind aus allen Teilnehmern gezogen. Der Gutschein wird dann an dich/Sie gesendet (via private message auf Twitter) nachdem die Daten gesammelt sind.

Verfahren

Um Aufschluss über Twitternutzung, deren Sprachwissenschaftlichen Implikationen, und die Persönlichkeitsmerkmale von Twitternutzern zu gewinnen, werden die Tweets von Studienteilnehmern gesammelt und von einem sprachwissenschaftlichen und einem persönlichkeitsanalytischem Fokus her analysiert.

Für die Datengewinnung bitten wir dich/Sie einen kurzen Onlinefragebogen auszufüllen, der Fragen zur Persönlichkeit (Big Five Faktoren), demographischen Informationen, Social-Media-Nutzung, und deinem/Ihrem Twitter handle (Twitternutzername, z.B. @primesty22) stellt. Dein/Ihr Twitter handle wird nur benötigt um deine/Ihre onehin öffentlich zugänglichen (!) Tweets zu sammeln.

Risko(s) oder besondere Belastung

Für Studienteilnehmer besteht kein(e) vorhersehbare(s) Risiko/Belastung. Für die Beantwortung des Fragebogens rechnen wir mit einem Zeitaufwand von maximal fünf Minuten.

Datenschutz/Vertraulichkeit

Wir weisen darauf hin, dass alle personenbezogenen Daten der Teilnehmer, sowie deren Tweets vertraulich behandelt werden (die Daten werden als kombiniertes Ergebnis verwendet ohne individuelle Nutzer hervorzuheben). Die Namen der Teilnehmer sind nicht notwendig für die erfolgreiche Analyse der Daten. Die Fragebögen werden verschlossen aufbewahrt und nur der Principal Investigator hat Zugriff auf die Daten. Die Daten werden auf der sicheren Qualtrics-Database gespeichert bis sie vom PI gelöscht werden. Da Qualtrics und die offene Twitter API für die Datengewinnung verwendet werden, verbleiben die Daten beim PI und werden zu keinem Zeitpunkt der Öffentlichkeit zugänglich gemacht!

Teilnahme

Die Teilnahme an dieser Studie ist vollkommen freiwillig! Du/Sie hast/haben das Recht jederzeit zurückzutreten oder die Teilnahme ganz zu verweigern.

Fragen zum Forschungsvorhaben

Falls du/Sie Fragen zum Forschungsvorhaben hast/haben, richte(n) (Sie) diese bitte an den Principal Investigator, Matthias Raess (mraess@bsu.edu).

Ethikkommission

Die vorliegende Studie wurde von der Ethikkommission der Ball State University genehmigt. IRB-Referenz-Nummer: # 979954-1

Fragen, die deine/Ihre Rechte als Studienteilnehmer betreffen, richte(n) (Sie) bitte an: Director of Research Integrity, Office of Research Integrity, Ball State University, Muncie, IN 47306, (765) 285-5070, irb@bsu.edu.

Ich habe die vorliegende Einverständniserklärung gelesen und verstanden und nehme aus freiem Willen an dieser Studie teil. Ja/Nein

****

Principal Investigator: Matthias Raess, PhD candidate, Ball State University, Department Phone: (765) 285-8580, Email: mraess@bsu.edu
Faculty Advisor: Dr. Carolyn MacKay, Professor, Department of English, Ball State University, Office Phone: (765) 285-8539, Department Phone: (765) 285-8580, Email: cjmackay@bsu.edu

Teil II – Big Five Fünf Faktoren Test (BFI-10)

In wie weit treffen die folgenden Aussagen auf Sie zu?

	Trifft überhaupt nicht zu	Trifft eher nicht zu	Weder noch	Eher zutreffend	Trifft voll und ganz zu
Ich bin eher zurückhaltend, reserviert.	1	2	3	4	5
Ich schenke anderen leicht Vertrauen, glaube an das gute im Menschen.	1	2	3	4	5
Ich bin bequem, neige zur Faulheit	1	2	3	4	5
Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	1	2	3	4	5
Ich habe nur wenig künstlerisches Interesse.	1	2	3	4	5
Ich gehe aus mir heraus, bin gesellig.	1	2	3	4	5
Ich neige dazu, andere zu kritisieren.	1	2	3	4	5

Ich erledige Aufgaben gründlich.	1	2	3	4	5
Ich werde leicht nervös und unsicher.	1	2	3	4	5
Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	1	2	3	4	5

(adapted from Ramstedt & John, 2007; Ramstedt et al., 2012)

Teil III – Demographische Informationen

Zum Abschluss würden wir gerne noch ein paar persönliche Angaben von Ihnen haben.

1. Bitte geben Sie Ihren Twitter-Nutzernamen ein (ohne @-Zeichen, z.B. (@)primesty22)

2. Nutzung von sozialen Medien (bitte wählen Sie alle aus, die Sie zusätzlich zu Twitter benutzen)

Facebook, Instagram, Flicker, LinkedIn/Xing, Tumblr, Reddit

3. Wie lange benutzen Sie Twitter schon?

4. Wie oft checken Sie ihre Twitter updates?

5. Wie viel Zeit verbringen Sie täglich auf Twitter?

6. Welche Funktion erfüllt der Hashtag auf Twitter?

Index-Suchfunktion auf Twitter, Ersatz für Text, Beides

7. Geschlecht? Männlich/weiblich

8. Alter

9. Bitte geben Sie die Postleitzahl ihres Heimatortes ein.

10. Ist das die gleiche Stadt in der Sie aufgewachsen sind?

11. Bitte geben Sie die Postleitzahl des Ortes ein, in dem Sie aufgewachsen sind.

12. Familienstand

13. Wie viel Geld haben Sie monatlich im Schnitt zur Verfügung?

14. Staatsangehörigkeit

15. Bitte geben Sie ihre Muttersprache an.

Deutsch, Englisch, Türkisch, Russisch, Französisch, Spanisch, Italienisch, andere

16. Schulabschluss

17. Beruf/Bildung

18. Erwerbstätigkeit

Appendix B: Questionnaire – English Version

Part I – Informed consent

Informed consent – German version
Project Title:
Interactions[3] – language, demographics, and personality; an in-depth analysis of German tweets

Welcome to the survey and thank you for your time!

Introduction

As part of a doctoral dissertation at Ball State University, we would like to ask people who actively use Twitter as a social medium to fill out a short online questionnaire.

The goal of the study is to establish a connection between Twitter usage, demographic information, and personality features of German Twitter users.

The survey takes about five minutes to complete.

Inclusion/Exclusion Criteria

We are looking for people who use Twitter actively (with a public profile) and are between the ages of 18 and 45.

Remember there are no (!) right or wrong responses; all responses are valuable to this research.

Incentive

As a thank you for participating in the study, you will have the chance to be drawn for one of three 20 EUR Amazon gift cards. Winners will be drawn at random and the gift card will be sent to you after data collection is completed (via private message through Twitter).

Procedure

To gain insight into Twitter usage, its linguistic implications, and the personality features of German Twitter users, your publically available tweets will be gathered and analyzed under a linguistic and personality research lens.

To do that, you are asked to fill out a short online questionnaire with a personality measure (the Big Five trait inventory), demographic data, your social media use, and your Twitter handle (user name, e.g. @primesty22). Your Twitter handle is only needed to collect your publically (!) available tweets from your Twitter timeline.
Risks/discomforts

There are no perceived risks for participating in the study. The total duration should not exceed 5 minutes.

Confidentiality

We would like to inform you that your data (personality measures, demographic information) as well as your publically available tweets are confidential and will be handled as such (the data will be presented as a combined result). Your actual name is not (!) needed for successful data analysis. The questionnaires will be stored on Qualtrics' secure server, to which only the principal investigator will have access (until they are deleted by the PI). As Qualtrics and the open Twitter API are used for data collection, the data will remain with the PI and will, at no time, be made accessible to the public at large.

Participation

Participation in this study is completely voluntary. You have the right to withdraw at any point or refuse participation all together.

Questions

If you have any questions about the study, please direct them to the principal investigator, Matthias Raess (mraess@bsu.edu).

Institutional Review Board

This research has been approved by the Ball State University Institutional Review Board.
IRB protocol number: # 979954-1

For questions about your rights as a research subject, please contact: Director of Research Integrity, Office of Research Integrity, Ball State University, Muncie, IN 47306, (765) 285-5070, irb@bsu.edu.

****

Principal Investigator: Matthias Raess, PhD candidate, Ball State University, Department Phone: (765) 285-8580, Email: mraess@bsu.edu
Faculty Advisor: Dr. Carolyn MacKay, Professor, Department of English, Ball State University, Office Phone: (765) 285-8539, Department Phone: (765) 285-8580, Email: cjmackay@bsu.edu

I have read and understood the informed consent form and desire of my own free will to participate in this study. Yes/no

Part II – Big Five personality inventory – BFI 10

How well do the following statements describe your personality? I see myself as someone who…

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
… is reserved	1	2	3	4	5
… is generally trusting.	1	2	3	4	5
… tends to be lazy	1	2	3	4	5
… is relaxed, handles stress well	1	2	3	4	5
… has few artistic interests	1	2	3	4	5
… is outgoing, sociable	1	2	3	4	5
… tends to find fault with others	1	2	3	4	5
… does a thorough job	1	2	3	4	5
…gets nervous easily	1	2	3	4	5
… has an active imagination	1	2	3	4	5

(adapted from Ramstedt & John, 2007; Ramstedt et al., 2012)

Part III – Demographic information

Finally, we would like to ask you a couple of personal questions.

1. Please fill in your Twitter handle (without the @-sign, e.g. (@)primesty22)

2. Please select all social media you use in addition to Twitter

Facebook, Instagram, Flicker, LinkedIn/Xing, Tumblr, Reddit

3. How long have you been using Twitter?

4. How often do you check your Twitter updates?

5. How much time do you usually spend on Twitter daily?

6. What is the function of the hashtag on Twitter?

Indexing (making tweets searchable), substitute for text, both

7. Gender – male/female

8. Age

9. Please fill in the ZIP code for your place of residence

10. Is this the same city you grew up in?

11. Please fill in the ZIP code for the city you grew up in

12. Relationship status

13. How much money do you have at your disposal monthly?

14. Citizenship

German/not German

15. What is your first language?

German, English, Turkish, Russian, French, Spanish, Italian, other

16. Education

17. Occupation status

Appendix C: LIWC – Tentative Words – German dict. 2001

German	English	Stem
angeblich	supposedly	*
beinahe	almost	
duerfte	might	*
eigentlich	actually	
erhofft	desired	*
erscheinen	appear	
erschien	appeared	
etwa	about	*
eventuell	maybe	*
fast	almost	
gehofft	hoped	*
gelegentlich	occasionally	*
gewettet	bet	*
gezaudert	hesitated	*
gezoegert	hesitated	*
gezweifelt	doubted	*
glueck	luck	*
hoffen	hope	*
hoffentlich	hopefully	*
hoffnung	hope	*
hoffnungen	hopes	
hoffnungslos	hopeless	*
hoffnungsvoll	hopeful	*
irgendei	something	*
irgendet	something	*
irgendj	someone	*
irgendwa	something	*
irgendwe	someone	*
irgendwie	somehow	
irgendwo	somewhere	*
jederzeit	anytime	*
jemand	someone	*
konflikt	conflict	
koennte	could	*
labil	unstable	*
manche	some	*
moeglich	possible	*
mutgemasst	assumed	*
mutmass	assumption	*
nahezu	almost	
oder	or	
probier	try	*

provisorisch	tentative	*
schaetz	estimate	*
scheint	appear	
scheintst	appear	
schien	appeared	*
sozusagen	as it were	
unbestimmt	indeterminate	*
undeutlich	indistinct	*
uneins	at a strife	
ungefaehr	roughly	*
ungewiss	uncertain	*
unklar	unclear	*
unsicher	uncertain	*
vage	vague	
vermein	alleged	*
vermut	suppose	*
verwirr	confuse	*
verworren	nebulous	*
vielleicht	maybe	
vorahnung	premonition	*
voraussichtlich	presumed	*
vorsichtig	careful	*
wahrscheinlich	likely	*
wette	bet	
wetten	bet	
wettete	bet	*
zauder	hesitate	*
zeitweise	occasionally	*
ziemlich	quite	*
zoeger	hesitate	*
zufaelle	coincidences	*
zufaellig	coincidentally	*
zufall	coincidence	*
zweifel	doubt	
zweifels	doubt	